

DOCUMENT RESUME

ED 222 178

IR 010 411

AUTHOR Goldin, Sarah E., Ed.; Thorndyke, Perry W., Ed.
TITLE Improving Team Performance: Proceedings of the Rand Team Performance Workshop.
INSTITUTION Rand Corp., Santa Monica, Calif.
SPONS AGENCY Office of Naval Research, Arlington, Va.
REPORT NO ISBN-0-8330-0249-X; R-2606-ONR
PUB DATE Aug 80
CONTRACT N00014-79-C-0753
NOTE 188p.; The proceedings of the Rand Workshop on Team Performance (Santa Monica, CA, November 27-29, 1979).

EDRS PRICE MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS *Decision Making; *Game Theory; *Group Dynamics; *Human Factors Engineering; Organizational Effectiveness; *Organizational Theories; Research Methodology; Research Needs; Simulation; Small Group Instruction; *Team Training

IDENTIFIERS Heuristic Models; Naval Training; Small Group Communication

ABSTRACT

The 16 papers in this collection discuss options and areas for future research on teams, including performance requirements, structure, communications, training techniques, and organizational determination of team performance, from the perspectives of varied disciplines; e.g., psychology, computer science, management science, and decision theory. These papers were presented at a workshop designed to review research in varied fields that might have potential relationships to team performance, consider promising directions for research, and promote discussion of potential interdisciplinary research methodologies and facilities for improving team training and research. The workshop was also intended to help identify current problems in training and Navy team performance. A summary of the major workshop themes is followed by the text of papers presented and summaries of follow-up sessions for the eight major workshop topics: gaming and simulation, organization theory, small group processes, cognitive psychology, training and instruction, heuristic modeling, decision theory, and human engineering. A 7-page bibliography is provided. (LMM)

* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *

ED222178

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)

✦ This document has been reproduced as
received from the person or organization
originating it.

Minor changes have been made to improve
reproduction quality.

• Points of view or opinions stated in this docu-
ment do not necessarily represent official NIE
position or policy.

R-2606-ONR
August 1980

Improving Team Performance: Proceedings of the Rand Team Performance Workshop

Sarah E. Goldin and Perry W. Thorndyke, Editors

A Report prepared for the
OFFICE OF NAVAL RESEARCH

"PERMISSION TO REPRODUCE THIS
MATERIAL IN MICROFICHE ONLY
HAS BEEN GRANTED BY

Tom Cockrell

TO THE EDUCATIONAL RESOURCES
INFORMATION CENTER (ERIC)."

Rand
SANTA MONICA, CA. 90406

APPROVED FOR PUBLIC RELEASE: DISTRIBUTION UNLIMITED

IR010411

PREFACE

Many military tasks require the skilled and coordinated effort of a team of individuals. Nevertheless, in a 1976 report prepared for the Secretary of Defense, the Defense Science Board Task Force on Training Technology concluded that "insufficient attention is now being given to collective training." As an indirect result of that study, the Psychological Sciences Division of the Office of Naval Research contracted with The Rand Corporation in 1979 to design a multiyear program of research on team training and effectiveness.

As part of that design effort, Rand sponsored a Team Performance Workshop in Santa Monica, California, on November 27-29, 1979. The purposes of the workshop were (1) to review the research in a variety of fields that potentially bear on various aspects of team performance, (2) to consider promising future directions for research on teams, and (3) to promote discussion of potential interdisciplinary research methodologies and facilities for improving team training and research.

The workshop comprised a set of presented papers by the sixteen invited participants and a set of discussion sessions including the participants, the workshop organizers, other Rand professionals associated with the design effort, and various Department of Defense personnel attending the workshop as observers.

The invited participants included cognitive psychologists, social psychologists, human factors psychologists, instructional psychologists, computer scientists, management scientists, and decision theorists. Participants were selected on the basis of recognized expertise within their respective disciplines, as well as their ability to broadly represent the issues and viewpoints characterizing each of these disciplines. Because the workshop was intended to encompass a wide variety of approaches and methods, speakers were encouraged to consider promising future directions rather than report on finished research results.

The workshop was organized around eight sessions, each devoted to one particular discipline area relevant to the study of team performance. These included gaming and simulation, organization theory, small group processes, cognitive psychology, training and instruction, heuristic modeling, decision theory, and human engineering. Within each session, relatively formal paper presentations were followed by an informal question and discussion period.

The organization of this report corresponds closely to that of the workshop. For each session, the texts of the presentations are followed by a summary of the major points raised in the subsequent discussion. An introductory section of the report provides an overview, outlining the major conceptual issues that emerged from the papers and discussions. This section includes much of the substance of the general discussions that opened and closed the workshop.

This report attempts to serve several purposes: to record the workshop proceedings, to summarize the major themes that emerged during the two and one-half days of meetings, and to suggest specific research directions for future research on teams. The interdisciplinary nature of the report should make it of interest to psychologists, sociologists, management scientists, organizational theorists, computer scientists, systems scientists, and military personnel concerned with the im-

provement of teams. In addition, it may assist R&D program managers in long-range planning for research on team performance.

In addition to this report, the Rand Team Performance Project has also produced a major report detailing recommendations for designing a large-scale team research effort. That work, R-2607-ONR, can be considered a companion volume to the present report, since it elaborates on many of the issues originally raised at the workshop.

The workshop was supported by the Office of Naval Research under Contract N00014-79-C-0753. Each of the Rand Team Performance Project staff members made important contributions to the success of the workshop: Polly Carpenter-Huffman, Theodore Donaldson, Larry Freeman, R. Stockton Gaines, Barbara Hayes-Roth, Frederick Hayes-Roth, Alain Lewis, Mark Menchik, Shelley Taylor, and Milton Weiner. Kay McKenzie and Mary Shannon retyped and edited the report manuscript.

SUMMARY

This paper reports the proceedings of the Rand Workshop on Team Performance sponsored by the Office of Naval Research. This workshop considered options for future research on teams from the perspectives of a number of different disciplines. Experts in the domains of gaming and simulation, organizational theory, small group processes, cognitive psychology, training and instruction, decision theory, heuristic modeling, and human engineering presented views on potential contributions of their disciplines to team research. Discussions among workshop participants focused on a range of issues, including current problems in the performance of Navy teams, the desirability of studying teams in their operational environments, the need for specifying training objectives prior to instructional design, and the need for designing team research programs prior to establishing a dedicated team research facility. A number of broad areas seem to offer promise for future research: team performance requirements, team structure, team communications, training techniques, and organizational determinants of team performance.

WORKSHOP PARTICIPANTS

Contributors

EUGENE BURNSTEIN, Department of Psychology and The Institute for Social Research, The University of Michigan
ALLAN COLLINS, Bolt, Beranek, & Newman, Inc.
JOHN PATRICK CRECINE, Dean, College of Humanities and Social Sciences, Carnegie-Mellon University
WARD EDWARDS, Social Science Research Institute, University of Southern California
CARL FEEHRER, Bolt, Beranek, & Newman, Inc.
SARAH GOLDIN, Information Sciences Department, The Rand Corporation
IRA GOLDSTEIN, Xerox Palo Alto Research Center
IRWIN GOLDSTEIN, Department of Psychology, University of Maryland
EARL HUNT, Department of Psychology, University of Washington
KENNETH MACCRIMMON, Faculty of Commerce and Business Administration, University of British Columbia
GARY MARTINS, Information Sciences Department, The Rand Corporation
DAVID MEISTER, Navy Personnel Research and Development Center
NILS NILSSON, SRI International
LARRY NOWELL, Logicon, Inc.
H. McILVAINE PARSONS, Institute for Behavioral Research, Inc.
WILLIAM RIZZO, Naval Training Equipment Center
KARLENE ROBERTS, School of Business, University of California, Berkeley
PERRY W. THORNDYKE, Information Sciences Department, The Rand Corporation

Other Workshop Participants

MARSHALL FARR, Office of Naval Research
LARRY FREEMAN, The Rand Corporation
R. STOCKTON GAINES, The Rand Corporation
LT. COL. ROBERT GIBSON, HQMC/OTOR 32
EUGENE GLOYE, Office of Naval Research
BARBARA HAYES-ROTH, The Rand Corporation
FREDERICK HAYES-ROTH, The Rand Corporation
JOHN KENNEDY, The Rand Corporation
BERT KING, Office of Naval Research
ELIZABETH LAMBERT, Naval Training and Equipment Center
JAMES LESTER, ONR Branch Office, Boston
ALAIN LEWIS, The Rand Corporation
WILLIAM LINDAHL, OASN, MRA&L
MARK D. MENCHIK, The Rand Corporation
DAVID STONNER, Office of Naval Research
SHELLEY TAYLOR, UCLA
MARTIN TOLCOTT, Office of Naval Research

MILTON WEINER, The Rand Corporation
ROBERT WESSON, The Rand Corporation
JOHN WINKLER, Harvard University

CONTENTS

PREFACE.....	iii
SUMMARY	v
WORKSHOP PARTICIPANTS.....	vii
Section	
I. AN OVERVIEW OF WORKSHOP ISSUES, by Sarah E. Goldin and Perry W. Thorndyke.....	1
II. GAMING AND SIMULATION	13
Navy Team Training: Some Critical Issues, by William A. Rizzo.....	15
A Look at Team Training, by Larry Nowell.....	26
Simulation-Based Team Training, by H. McIlvaine Parsons	34
Discussion of Session 1	50
III. ORGANIZATION THEORY.....	51
Human Information Processing Approaches to Organization and the Study of Teams: Possible Research Directions, by John P. Crecine.....	53
Some Thoughts About Effective Team-Effectiveness Research, by Karlene H. Roberts	60
Discussion of Session 2	66
IV. SMALL GROUP PROCESSES.....	69
Processing Arguments and Forming Opinions in Groups, by Eugene Burnstein	71
Discussion of Session 3	86
V. COGNITIVE PSYCHOLOGY.....	89
A Cognitive Science Approach to Research on Teams, by Allan Collins.....	91
A Cognitive Science and Psychometric Approach to Team Performance, by Earl Hunt	94
Discussion of Session 4	108
VI. TRAINING AND INSTRUCTION	109
An Answer to Team Training Instructional Techniques: The Design of Needs Assessment and Evaluation Procedures, by Irwin L. Goldstein	111
From Personal to Cooperative Computing, by Ira Goldstein	118
Discussion of Session 5	122

VII.	HEURISTIC MODELING	123
	Teams and Distributed Artificial Intelligence, by Nils J. Nilsson	125
	The Future of Heuristic Modeling in Team Training Applications, by Gary Martins	129
	Discussion of Session 6	137
VIII.	DECISION THEORY	139
	In Decisionmaking, Are N Heads Better Than One?, by Ward Edwards	141
	Studying Team Performance: A Decision Theory Perspective, by Kenneth R. MacCrimmon	144
	Discussion of Session 7	159
IX.	HUMAN ENGINEERING	161
	Human Factors In Teleconferencing, by Carl Feehrer	163
	The Role of Human Factors Engineering in Team Performance Research, by David Meister	176
	Discussion of Session 8	183
	BIBLIOGRAPHY	185

I. AN OVERVIEW OF WORKSHOP ISSUES

Sarah E. Goldin and Perry W. Thorndyke

The papers presented during this workshop represent a variety of approaches to team research design. Each author discusses a number of important team research issues from the perspective of his or her own discipline. However, no single paper presents a broad, integrated view of the research issues.

Therefore, we have attempted to develop an overview of the workshop from an interdisciplinary perspective. The following discussion surveys the major themes that emerged during the workshop sessions and provides an integrating framework for the observations and recommendations of individual contributors. This integration highlights several conclusions representing the consensus of workshop participants:

1. An investigation of the dimensions of teams and team performance should precede particular team research efforts.
2. Current team training evaluation and deployment practices suffer from a number of serious deficiencies.
3. These deficiencies suggest several principles to guide future team research.
4. Several promising research areas could address these deficiencies immediately and directly.

The remainder of this overview elaborates on each of these themes. The first section summarizes important characteristics of teams and team functioning alluded to throughout the workshop. The second section synthesizes the participants' views of the major problems afflicting current Navy teams. The third section discusses four principles that many workshop participants suggested as guides to future team research. The final section surveys some interdisciplinary approaches to team research endorsed by many of the participants.

THE NATURE OF TEAMS

A pervasive concern of workshop participants is the requirement for a taxonomy of teams based on their critical attributes. Such a taxonomy would include categories of team tasks and dimensions of variation distinguishing different types of teams.

The development of a team taxonomy as the first step in a research program would serve several functions. First of all, such a taxonomy could guarantee the utility of the research results to actual Navy teams. Much of the prior work on team performance is of limited use to the Navy because of substantial differences between the studied teams and actual Navy teams. In particular, many laboratory studies of teams fail to simulate features of the actual task environment that may critically affect team performance. A team taxonomy would enable researchers to identify and preserve these critical features in the research environment.

Second, a taxonomy of teams would facilitate the isolation of potential variables for study. The analysis of teams according to a set of structural and operational features would encourage the design of research focusing on a few critical variables while controlling other potentially confounding variables.

Finally, a taxonomy would enable researchers to judge the appropriate level of abstraction to which their research could be generalized. The taxonomy would permit the assessment of similarity among teams and the aggregation of teams into equivalence classes that share functional attributes. Such classifications would also allow researchers to choose prototypical or representative teams as vehicles for research.

A number of workshop participants suggest the rudiments of such a taxonomy. The papers of Rizzo, Meister, and MacCrimmon offer explicit definitions of teams, while others suggest characteristics distinguishing teams from other groups. Three characteristics seem particularly critical in distinguishing among types of teams and team functions: team procedures, structure, and degree of coordination. We treat each of these areas separately below.

Team Performance of Procedures

Teams exist as vehicles to perform particular tasks or achieve particular goals. As Meister points out, the nature of the task procedures and equipment may determine the team requirements for the task. For example, a P-3C aircraft requires a crew comprising a pilot, a co-pilot, and a flight engineer (see Rizzo, pp. 15-25). Each team member performs particular functions according to the equipment he must operate and the current goals of the team.

The rigidity and formality of these procedures vary widely across teams. Edwards, Hunt, and other workshop participants note that Navy team tasks fall into two broad categories: perceptual/motor tasks (e.g., gun loading) and information-processing tasks (e.g., tactical decisionmaking). Perceptual/motor tasks generally require a fixed set of procedures that can frequently be formalized in an equipment manual. Information-processing tasks, in contrast, often lack clear procedural definition. To some extent, the degree of procedural rigidity depends on the predictability of the environment in which the team must function. Tasks that can be anticipated and specified in advance typically have established procedures. On the other hand, tasks with wide variability in initial conditions and potential outcomes require procedural flexibility.

The degree of procedural rigidity impacts on the measurement of team effectiveness. For highly routinized teams, team effectiveness depends on the correct execution of team procedures by each team member. However, the effectiveness of teams with poorly specified procedures and goals is more difficult to assess, because of the absence of specific performance criteria and principles for controlled observation.

Team Structure

Team structure can be defined as the pattern of relationships or dependencies that exist among team members. The structure of teams can vary in several ways. One dimension of variation is defined by the potential interactions within the team.

In a hierarchical structure (exemplified by the standard chain of command), each team member interacts with a limited number of other members—specifically, those immediately above and below him in the hierarchy. In contrast, a linear or “flat” organization distributes authority across team members and permits more extensive intrateam interaction. Another dimension of variability is the rigidity of team structure. Some teams have a fixed set of roles and relationships for all situations. Other teams have a number of structural configurations that vary with task characteristics or situational constraints. A P-3C flight crew exemplifies the former type of structure, while a Combat Information Center (CIC) team illustrates the latter.

These structural dimensions are not completely independent. In general, strict hierarchical structures that preserve the military chain of command tend to be the least flexible. In such organizations, military rank often determines the pattern and content of team interactions.

Collins, Hunt, Roberts, Nilsson, and Crecine all consider possible relationships between team structure and performance. One popular hypothesis suggests that certain structures are optimal for certain types of tasks. The optimal structure for a team may depend on the nature of the various subtasks, their sequential dependencies, and the team interaction requirements. Crecine suggests that an appropriately chosen team structure can compensate for the memory capacity and serial processing limitations of individual human information processors. Teams can perform several tasks in parallel, and individual team members can specialize in monitoring and retaining particular types of information. However, Crecine also hypothesizes that highly structured teams may be more restricted than individuals in the breadth of tasks they can perform. Thus, structured teams may trade off performance flexibility for processing power.

Team Communication and Coordination

A third important feature of teams is their reliance on communication and coordination. To the extent that team performance relies on the interaction among team members, methods for developing effective communications techniques and coordination skills become important research goals.

Several aspects of communication and coordination seem critical in differentiating various types of teams. One distinguishing characteristic centers around the technology of communication. The papers by Parsons, Feehrer, and Ira Goldstein illustrate this range of potential communications requirements. Another aspect of team interaction is the degree of specificity of communication protocols. From a research perspective, the more interesting teams may be those that have loosely defined communications networks and protocols. Such teams may develop special jargon that enables team members to convey information rapidly and efficiently. However, these evolutionary changes in team communications patterns could interfere with team functioning in situations where team membership is not stable over time.

Collins and Martins, in particular, emphasize that team communication extends far beyond the objective transmission of information. Communication engages a whole range of cognitive processes, from memory retrieval to inferential reasoning. Since communication strategies dictate that what is already known can

be omitted from the explicit messages, both the person transmitting the information and the person receiving it must constantly make inferences about the other's knowledge and goals. The pressures toward conciseness and accuracy in military team communications are likely to accentuate these cognitive demands.

PROBLEMS IN CURRENT TRAINING AND PERFORMANCE OF NAVY TEAMS

Many of the papers and discussion comments focus attention on the identification and clarification of current problems affecting Navy teams. Contributors from the areas of operational simulation, organization theory, instruction, and human engineering all note deficiencies in organizational policies regarding the formation, training, and maintenance of effective teams. These deficiencies fall into two broad categories: organizational policies that constrain the function of standing teams and policies that govern training practices.

Organizational Constraints on Team Effectiveness

The organizational context in which Navy teams function may seriously limit their effectiveness. Many problems affecting teams stem from policies that determine composition, turbulence, and motivation within existing teams. For example, the incentive structure of the Navy may undermine the effectiveness of many teams. In general, salary and promotions are allocated to individuals rather than teams. A team member may thus strive to maximize personal performance rather than the overall performance of his team. The policy of selective, merit-based promotion of individuals also tends to increase team turbulence. Frequent personnel changes reduce the extent to which team members can develop specialized strategies for coordination. In addition, promoting an effective team member out of the team leaves less effective team members to carry on team functions.

Economic and pragmatic considerations may also influence team effectiveness by dictating decisions concerning the amount and type of team training administered. For example, the number of individuals who can be trained and the extent of training each receives may depend on the availability of simulators and instructors, and the amount of time allocated to training exercises. Current manpower shortages make any training at all a costly proposition because training programs reduce the number of personnel in a state of operational readiness.

Organizational attitudes can create further barriers to improving team effectiveness. Operational and training commanders may be reluctant to introduce technological innovations that promise improvements in team performance. A major factor in this reluctance may be the high overhead, in time and expense, associated with instituting major changes. The lead time required to select, install, shake down, and evaluate new systems may frequently exceed the duration of the decisionmaker's command, thus providing a disincentive to undertake innovation.

A final problem centers on the frequent lack of standardization. Current doctrine treats teams as systems of interchangeable parts (human and machine). However, equivalent teams at different sites may operate different equipment and use different procedures. Transfer from one installation to another may be extremely

difficult, since previous experience may hinder rather than assist the learning of slightly altered skills.

Problems with Team Training Practices

In addition to discussing organizational constraints that limit the effectiveness of teams and team training, workshop participants also note a number of problems with the training practices themselves. Most of these problems reflect an ad hoc approach to team training and evaluation without systematic attention to instruction design.

Rizzo argues that the most pervasive problem with current Navy team training is its lack of clearly defined instructional objectives. The goals of training are typically stated in very general terms (e.g., "improved operational skills"), rather than in terms of specific procedural skills or performance criteria. Various participants note that team training has two different goals: to promote the development of procedures for accomplishing task performance and to improve coordination among procedures performed by different team members. Although the training methods required to achieve each of these goals may differ greatly, most Navy team training programs do not recognize the distinction. In fact, much team training is devoted to improving individual task skills, a function that may be best accomplished outside of the team training program.

Another training problem related to lack of instructional objectives is the lack of objective evaluation criteria. Gibson points out that the two problems are closely related because evaluation criteria inevitably become training objectives. That is, tests or other evaluation instruments determine the content and emphasis of the training programs they are designed to assess. Currently, most evaluation during training relies on the instructors' subjective judgments of individual and overall performance. These measures provide little insight into the details of team functioning. Ideally, evaluation should indicate exactly where and how the performance of component subtasks should be improved.

Rizzo notes that the lack of clear evaluation criteria is compounded by the lack of consistent, scientific methods for providing feedback. Performance cannot improve without some knowledge of results. However, without effective evaluation measures, feedback may be vague, subjective, and of limited utility as a learning aid.

Lack of standardization in training equipment and curriculum content is a fourth major problem with Navy team training. Nowell notes that simulators at different sites may have different layouts, different consoles and displays, and different executive programs. According to Rizzo, the scenario exercises used in training are non-standard, often developed and modified as the exercise proceeds. The lack of specific training objectives makes it difficult to compare alternative curricula in terms of their training effectiveness.

Inconsistencies between the training situation and the operational environment pose another serious problem. Simulators that differ substantially from operational equipment, and unrealistic or oversimplified scenarios, tend to promote negative transfer between the training and operational environments.

A final problem with Navy team training practices concerns the relative emphasis given to team versus individual training. Because training programs lack clear-

ly stated instructional objectives and detailed performance measures, they do not distinguish skills best trained at the individual level from those requiring team training. Much current team training develops job skills that could readily be trained on an individual basis. This suggests the need for an articulated theory of teams that would include principles for distinguishing when team, as opposed to individual, training is appropriate.

SOME GUIDING PRINCIPLES FOR FUTURE TEAM RESEARCH

In considering options for future research on Navy teams, the workshop as a whole focused on developing a viable framework in which to conduct the research. Motivated by this concern, workshop participants proposed a number of broad guidelines for future team research. This section presents three guidelines that received substantial support during workshop discussions.

1. Study Teams in Situ

Research of potential utility to the operational Navy requires a background understanding of the functions, constraints, and environment of Navy teams. For this reason, many workshop participants recommend studying Navy teams, their equipment, and their procedures in the operational setting. One compelling justification for studying teams in situ is the limited practical utility of previous laboratory studies of teams. Further, the need for a team taxonomy should motivate observation within the operational environment, since the precision of any derived taxonomy will depend on sampling a representative group of task environments and team behaviors.

In situ research, however, does not necessarily imply "applied" research. Parsons, Roberts, Frederick Hayes-Roth, and other participants argue that the classification of research as "basic" or "applied" depends on the research goals and the questions under investigation, not on the setting in which the research is conducted. They note that the research initiatives discussed during the workshop qualify as basic rather than applied studies because of the emphasis on hypothesis generation and testing rather than on optimization and control. In fact, such research would provide a foundation of methods, knowledge, and theories to support more applied studies of training and performance.

An emphasis on in situ research should also not be construed as a total rejection of laboratory approaches to studying teams. Many questions raised during the workshop can be investigated only within a laboratory setting. The first phase of these laboratory investigations, however, should focus on observational studies of actual teams in their operational environment. This approach will increase the likelihood that laboratory settings for team research accurately represent relevant aspects of the task environment that impact on functioning teams.

2. Analyze the Task Prior to Instructional Design or Implementation

A number of workshop participants caution against a recurring fault in instructional design: the design of an instructional program around available methods or technologies rather than around a set of training objectives and evaluation criteria.

This "cart before the horse" strategy can result in expensive and sophisticated instructional systems that fail to produce effective team members. Irwin Goldstein, Parsons, and Meister all recommend a multistage plan for the design of an instructional program that has equipment design and implementation as the final stages in the design sequence. These stages should always be preceded by extensive analysis of the task environment and careful formulation of training objectives and evaluation standards.

The first stage in the sequence, task analysis, requires a specification of the component tasks and skills to be trained. For team tasks, these components include the roles (i.e., the personnel) required by the task, the individual subtasks assigned to each role, the interaction among subtasks, and the sequencing of subtasks. Such an analysis should produce a detailed understanding of the cognitive, organizational, and material resource requirements for the task.

The second stage in the design sequence entails the formulation of training objectives based on the analysis of task components. A set of realistic evaluation criteria should also be established at this step in the design process. Evaluation criteria should be matched to instructional objectives, so that the rate of acquiring component skills can be reliably assessed.

During the next phase in the design sequence, training objectives are translated into concrete training procedures. The instructional program should incorporate opportunities for learner evaluation and feedback. At this point in the design sequence, available training technologies will obviously impact the selection of particular training procedures. However, such implementation decisions will be responsive to previous design decisions, driven by research needs rather than by available or popular technologies.

3. Design Team Research Facilities around a Research Program

The third guideline for team research represents an application of the above design principles to the entire team research program. One mechanism for conducting future team research is through the establishment of one or more sites as dedicated facilities for team research. Parsons in particular warns against considering the characteristics of such a facility before specifying research program objectives. In order for research facilities to be effective, design decisions about facilities must be secondary to decisions about the research they will support.

For example, the selection of a team to be studied in a simulated environment may introduce space, hardware, and data collection requirements. Ideally, laboratory equipment will simulate, functionally if not physically, the corresponding operational equipment. Early design of experimental tasks and methods may facilitate the automation of data collection and analysis. In general, the design of a facility around the program will result in a set of research tools tailored to the needs of the research to be conducted there.

SOME REPRESENTATIVE AREAS FOR INTERDISCIPLINARY RESEARCH

The workshop papers and discussions suggest an impressive number and variety of areas for future research on teams. While many topics derived from the

research traditions of a single discipline, several topics attracted the interest of many of the workshop participants. This section briefly considers areas that received the most extensive discussion over the course of the workshop.

Procedural Flexibility

Teams that perform in complex, dynamic environments must often respond to surprising, unexpected stimuli. For such teams, it is unlikely that trained procedures are sufficient to handle all these emergent situations. This raises a vital research question: How can flexibility in team procedures be exploited to improve overall team performance?

The workshop discussions contain lively debates on the issue of flexibility. Weiner, Kennedy, Tolcott, and others advocate a measure of procedural flexibility for teams. They suggest that strictly defined procedures reduce the team's adaptability in unusual situations and fail to accommodate strengths and weaknesses of individual team members. Bert King defends a stricter specification of standard operating procedures (SOPs), arguing that overly flexible procedures would not guarantee the highly reliable and automatic performance required in many combat environments.

To address these different perspectives, research on flexibility might examine the environmental conditions that favor either flexibility or procedural rigidity. One reasonable hypothesis is that flexible procedures may be desirable in highly unpredictable environments. Flexible procedures may also be desirable in benign environments with minimal time pressure, since they may improve team motivation and morale. On the other hand, in threatening situations with considerable time stress, rapidly executable SOPs may be necessary to insure the teams' survival.

Related research could also examine alternative methods for achieving procedural flexibility. One potential method might involve the design of flexible SOPs. That is, it may be possible to design procedures that incorporate a variety of options or strategies to handle diverse situations. Obviously, this technique will work only for tasks that can be performed in several ways and for which most situations can be predicted in advance. Another approach to flexibility would be "developmental": In a reinforcing environment, individual teams would be allowed to evolve idiosyncratic (but consistent) procedures best suited to the styles and skills of individual team members. This approach seems best suited to long-standing teams with low turbulence. The developmental approach has a major disadvantage, namely that procedural evolution may require a considerable time period. However, this method may offer opportunities for significant improvements in team coordination and morale.

Team Structure

One of the common distinctions between teams and non-team groups is the relatively well-defined role structure of teams. Historically, structure has been one of the primary foci for research on teams. The workshop papers and discussions include many hypotheses about the influence of structure on team performance. Two issues in particular illustrate the potential scope of research on team structure: the relationship between optimal structure and task characteristics, and the effects

of informal structure and leadership on performance. These topics are treated below.

Many participants suggest that the optimal team structure depends on the nature of the task. Hunt's paper elaborates this hypothesis in detail, drawing an analogy between two types of team organization and two popular control structures used in artificial intelligence programs. One control structure is hierarchical with authority located at the top of this hierarchy. It includes a single, powerful executive component that selects actions or "operators" capable of reducing the difference between the current and desired state. This centralized control structure performs most effectively in tasks with a single, well-defined goal and a limited set of potential operators. In contrast, the second type of structure exploits decentralized decisionmaking. In this organization, operators actively monitor the state of the environment and "trigger" or recommend themselves when they are appropriate to handle current conditions. Most of the knowledge concerning the appropriateness of particular actions resides in the operators themselves. The executive merely resolves conflicts between competing operators. This decentralized control structure performs most effectively in tasks with many poorly specified goals that must be achieved simultaneously using a large and diverse set of potential actions.

Hunt likens the centralized control structure to a hierarchical team in which the team leader makes all relevant decisions and relies on team members to carry them out. In contrast, the decentralized control structure resembles a committee structure in which team members cooperate on decisions and the team leader acts as a coordinator. Hunt proposes that each of these team structures could optimize team performance if matched to tasks with characteristics appropriate to that control structure. Research designed to test this and similar hypotheses offers a high potential payoff.

The relationship between formal and informal team structures represents another promising research area. Teams may often have a dual structure: a formal structure in which authority depends on rank or official designation, and an informal structure in which leadership depends on specialized expertise or interpersonal skills. Considerable research on teams in other organizations has investigated various aspects of this dichotomy. Similar studies of Navy teams might provide new insights into the dynamics of intrateam interactions.

One question that immediately arises concerns the evolution of informal structures. Under what conditions do informal structures emerge? One hypothesis suggests that such structures emerge when the formal structure fails to accommodate the requirements of the team task. If this were the case, then it would be interesting to determine whether there are differences between different kinds of teams in their propensity to evolve informal leadership structures.

A related research question concerns how informal structures affect overall team performance. In some situations, the informal structure might enhance performance (e.g., if the informal leader were more knowledgeable or skilled than the formal leader, or if the formal structure were inappropriate for the task). In other situations the informal structure might undermine performance (e.g., if serious conflicts arose between the formal and informal leaders). When an informal structure does enhance team effectiveness, perhaps that structure should receive official support, or alternatively, the formal structure of the team should be reorganized.

Team Communication

Communication, like structure, is a topic area addressed by almost every contributor. Of the many hypotheses and research proposals raised during the workshop, most fell into one of three areas: research on the technology of team communication, research on semantic aspects of communication, or research on pragmatic aspects of communication.

The technological options for team communication systems offer numerous research opportunities. Feehrer's studies of teleconferencing and Ira Goldstein's discussion of cooperative computing environments illustrate the range of potential research issues. Both papers point out that the primary design problem lies not in communications hardware, but in the design of systems tailored to the needs and capabilities of human information processors. Thus, research might investigate the information and communication requirements of specific teams and the interaction between these requirements and available technologies.

Clearly, the study of team communication requires the consideration of the content, or semantics, of the interactions. Martins argues that the central problem in team communication is not how to make wires survive longer, but how to convey meaning completely yet concisely. Collins' paper suggests two approaches to the analysis of communications content and structure. The first, called continuation analysis, classifies utterances according to the role they play in the overall communication protocol. This analysis might lead to prescriptions for optimal types of exchanges in particular situations. The second approach entails the analysis of indirect speech acts, or implicit presuppositions, underlying conversational exchanges. This kind of analysis might illuminate some of the sources of misunderstanding or failed communication within teams.

The pragmatics of team communication refers to the goals and intentions of various speakers and how they are understood and translated into action. Nilsson's paper on robot teams illustrates that getting someone else to do something by telling him about it is far from a simple problem. Collins argues that the lack of shared goals among team members may cause performance decrements. The study of communication pragmatics subsumes linguistic and psychological issues such as discovering implicit goals and intentions, modeling the knowledge of other team members, and using conventional structures for conversation. Research addressed to these issues may enable us to design communication protocols that help team members to better identify each other's goals and intentions.

Team Training

Team training methods and principles constitute a fourth area that workshop participants single out for research attention. Many of the research issues in this area arise from a consideration of problems in current training practices. Representative issues include the administration of feedback, simulator realism, and instruction in general teamwork skills.

As noted earlier, feedback in Navy team training tends to be administered erratically and unsystematically. Research is needed to ascertain the most effective practices for administering feedback in team training environments. Such research might investigate several attributes of potential feedback, including the recipient of the feedback (individuals or entire teams), the content (information on process

or information on outcomes), the timing (during or after the exercise), and the frequency.

Another training issue concerns the realism of training simulators. Training equipment often differs substantially from the corresponding operational equipment. Research is needed to determine how much these differences influence the effectiveness of training. Edwards, Parsons, and Kennedy suggest that requirements for highly realistic training equipment vary with the objectives of the training program. A program designed to train information-processing and decision-making skills probably requires a physically less realistic simulation environment than a program designed to train perceptual discrimination or motor responses. If this hypothesis is correct, then it might be possible to train certain teams economically in schematic environments, eliminating the need for expensive simulators of the classical kind, or other equipment.

A third recurring topic of discussion in the area of training research concerns the identification of general teamwork skills. Thorndyke, Weiner, and other participants suggest that skills for coordination and team problem-solving, robust over a variety of team tasks, may be identifiable and trainable. Edwards suggests that such skills might be identified through intensive study of "gold-plated teams"—that is, teams with efficient, near-optimal performance. This inductive method may suggest hypothetical skills or processes to be experimentally investigated. If such skills can be identified, additional research could seek methods for developing these skills through training.

Organizational Determinants of Team Performance

Many factors external to the team may significantly influence the team's effectiveness. Such factors may be referred to under the general heading of organizational determinants of team effectiveness. The workshop discussions raise a number of questions for potential research in this area. The problems cluster into three categories: team design policies, team turbulence, and organizational priorities.

A major limiting factor in team effectiveness may be in appropriate teamwork design. Teamwork design entails the determination of the people, procedures, and equipment required to perform a task. Meister's paper argues that teamwork design should proceed in a carefully monitored sequence of stages. At each stage, the designer should pose and answer a set of critical questions. For example, the first such question is, Do we need a *team* to perform this task? Meister suggests that many problems with current teams stem from faulty design. He recommends studying the design of past Navy systems requiring teams and reconstructing the sequence of decisions leading to the final designs. Such an historical reconstruction would provide the basis for identifying and correcting flaws in the design process, and possibly for tracing these flaws to assumptions in the Navy design philosophy.

Another factor limiting performance may be team turbulence. Turbulence refers to the turnover in team members due to rotation, attrition, or infirmity. Military incentive structures may reduce the effectiveness of teams by unintentionally encouraging team turbulence. In general, the military evaluates and promotes individuals rather than intact teams, basing promotional decisions on past performance. Such policies guarantee high turnover in team personnel and, in particular, result in the most proficient team members being removed from the team and replaced by less competent members. Routine rotation of team members further

exacerbates the potential negative effect of turbulence. Experimental research might evaluate the magnitude of such effects on performance. In addition, policy studies might explore alternative personnel policies that would enhance rather than hinder team performance. For example, in some cases it may be feasible to reward entire teams rather than individuals.

Finally, the utility of research results for improving team effectiveness depends in part on the ability and willingness of the Navy to implement the findings in operational settings. This introduces a set of policy questions that may ultimately determine the nature and success of the research program. These questions include: What do operational personnel and policymakers perceive as the most serious team problems? How flexible are Navy policies on promotion (or team assignment, task restructuring, training procedures)? How do training commands assess the costs and benefits of proposed innovations in training methods? Answers to such questions could guide the design of team research so that it addresses problems that are the most critical to the Navy and also have scientific merit. Such studies would also indicate the problem areas where research results would be most readily accepted and implemented.

SUMMARY

In this overview, we have attempted to synthesize the issues and proposals discussed by individual workshop participants into a more general framework for considering future team research. First, we considered how some of the defining characteristics of teams might relate to team effectiveness. Next, we surveyed a few of the current problems facing Navy teams, problems that stem from organizational constraints and a lack of theory-based training. Third, we presented some guidelines for future team research, drawn from the workshop papers and discussions. Finally, we discussed the five specific content areas most frequently recommended by workshop participants for future research. These included team performance requirements, team structure, team communication, training techniques, and organizational determinants of team performance. In each area we presented representative examples of proposed research topics and hypotheses.

Future studies of team performance have enormous potential. The breadth of team types and the variety of research proposals considered during the workshop testify to this fact. In addition, there exists a community of researchers capable of conducting these studies. Each of the diverse disciplines represented at the workshop can contribute to an integrated understanding of teams and how they function. In fact, team research may suffer from an embarrassment of methodological and conceptual riches. Obviously, a program of limited resources can sample only a few of the research issues and problem domains open to investigation. Focus and selection will be primary problems in planning future research. While the workshop was not intended to provide a plan for Navy team research, it does delineate key problem areas and potential research methods that can serve as input to team research design. Outlining such considerations is not sufficient to generate a research plan; it is, however, a necessary first step.

II. GAMING AND SIMULATION

NAVY TEAM TRAINING: SOME CRITICAL ISSUES

William A. Rizzo

INTRODUCTION

The interdependence of human behavior is a prevailing attribute of Navy operations. As a result, considerable resources are committed to team or crew training activities. Training not conducted at sea is commonly accomplished using land-based representations of operational platforms, having varying degrees of operational and physical realism. This paper will be concerned with these more contrived, land-based training environments and practices.

The justification for team training appears straightforward in terms of face validity, i.e., it seems logical that if the operational environment is modeled, and crews receive practice in war game scenarios, effective crew training will result. In principle, this is compelling logic; however, when examined from a human behavior point of view, the efficiency of these practices is suspect. I would like to explore this thought further by describing, in general, existing team training practices and shortcomings and illustrating a cross section of examples. I will describe what I feel are some of the key issues for future research and, finally, suggest a potential framework for the application of our efforts in improving team performance.

NAVY TEAM TRAINING PRACTICES

Formal tactical team training is conducted in two basic crew composition configurations: (1) school-assembled crews, and (2) intact operational crews. In a typical initial or transition training school, individuals are either arbitrarily assigned to crews based on number constraints (i.e., small class size) or crews are assembled based on an "optimal" mix of strong and weak students. Team training organizations concerned with refresher or proficiency training characteristically receive intact operational crews.

With the exception of certain types of refresher training, team training normally consists of complementary classroom, simulator, and in some cases, underway training. Simulator training typically consists of exercising a given crew, or some combination of crews, in a series of scenarios of graded difficulty. The scenarios are designed to model tactical situations which might be encountered in an operational environment. Some attempt is normally made in the initial phase of training to establish the relative proficiency of a given crew vis-à-vis the scenario exercises available. From this initial assessment of crew proficiency, a start point is determined for the difficulty level (or, perhaps more accurately, the complexity level) of subsequent exercises.

Much of existing team training practice represents a significant departure from contemporary educational technology. It is difficult to determine objectively just what is accomplished by team training, and it is apparent that much could be done to improve the team training process. Naval training personnel, for the most part,

apparently do not possess a clear understanding of what specific objectives should be accomplished by team training. Training objectives are stated in loose, general ways not optimally conducive to the development of well-ordered training programs. Most of the training programs I have examined fail to reflect the application of the systematic procedures of educational technology for training program development. Training is conducted largely through the medium of practice with non-standard but structured exercises. Deliberate systematic application of feedback for trainee guidance and error correction is noticeably deficient in most cases. Clearly stated, objective criteria and procedures for evaluating team performance are apparently not available.

NAVY TEAM TRAINING DEVICES

I would like to illustrate a somewhat representative cross section of the vehicles employed for the training of teams. These examples are not necessarily correlated with my earlier comments critical of training practices. Descriptions of these trainers will be deliberately brief; however, some of these, and other trainers are described in detail in Hall and Rizzo (1975).

The Multi-Class Advanced Submerged Control Trainer, Device 217C (Fig. 1) is designed to provide generic training in steering, diving, and casualty control operations for a variety of Fast Attack and Fleet Ballistic Class submarines. The principal components are:

- Steering and diving station
- Ballast control station
- Motion platform
- Instructor station
- Computer
- Communications system

The Submarine Combat System Trainer, Device 21A37/4 provides training in offensive and defensive tactics for nuclear submarine attack center crews. The trainer consists of three simulated submarine attack centers which may be operated independently or in coordination. The attack center mockups represent several classes of nuclear submarines with the following major equipment groups:

- Fire control systems
- Navigation and plotting equipment
- Target detection equipment
- Communications systems

The 1200 PSI Propulsion Plant Trainer, Device 19E22, provides indoctrination and training in engineering operating and casualty control procedures, systems tracing, watch standing, automatic boiler control operation, and inspection for Frigate Class ship engineering spaces. The trainer consists of a full-scale operational simulation of:

- Auxiliary machinery room No. 1 and No. 2
- Engine room

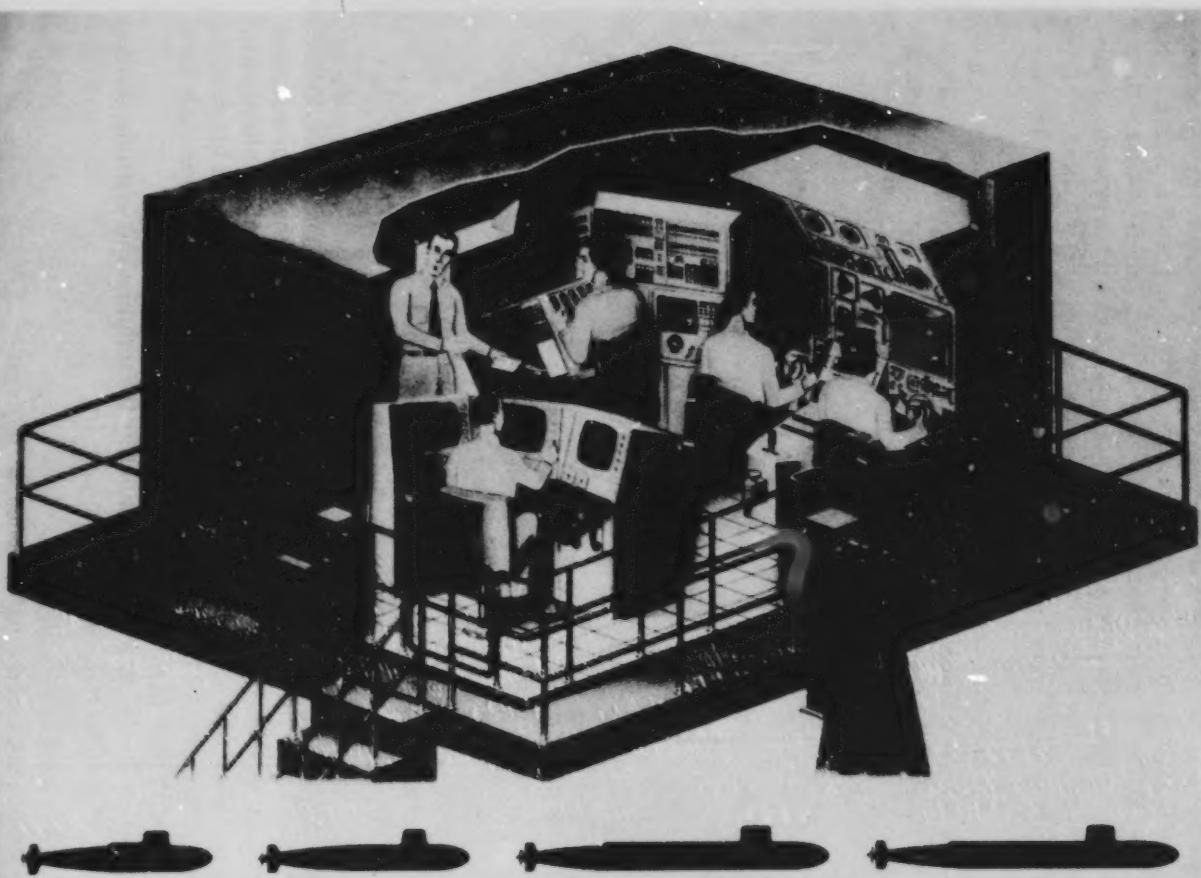


Fig. 1—Multi-class advanced ship control trainer device 21C7

- Fire room
- Forced draft blower room
- Electrical central

These spaces may be operated independently as separate trainers or in conjunction, for coordinated training of all engineering functions.

The Surface Ship Anti-Submarine Warfare (ASW) Early Attack Weapons Systems Team Trainer, Device 14A2 (Fig. 2), is used to train ship crews in the proper utilization of operational ASW systems. Training emphasizes the procedural, tactical decisionmaking, and crew coordination activities in employing these ASW systems. The trainer is also used in developing and planning advanced naval undersea defense tactics.

The trainer duplicates the physical configuration of major operational compartments and equipment of surface ship ASW attack weapons and simulates their functional operation and responses such as target detection, fire control solution, and weapon launching and tracking. The trainer occupies over 3000 feet of floor space, divided into six operating areas:

- Underwater battery plot—includes sonar, fire control, and communications equipment.
- Combat information center—for collection, evaluation, dissemination, and display of own ship tactical situation.
- Launch captain's central station—for status and control of weapons launching.
- Conning station—for Deck Officer ship control and monitoring of weapons status.
- Computer and projection equipment room.
- Problem critique and display room—for problem control, display, and debriefing.

The ASW Coordinated Tactics Trainer, Device 14A6 (Fig. 3) is designed to train decisionmaking personnel in the tasks they must perform when engaged in coordinated ASW tactics. Simultaneous operation of 48 vehicles of various types and numerous sensors can be simulated. Communications facilities simulate the various radio channels employed operationally to coordinate all phases of an ASW mission from search through attack. This trainer provides a synthetic environment within which personnel can practice collecting and evaluating ASW information, making decisions, and implementing the decisions. The trainer is not intended to train equipment operators; therefore, simulated equipment does not resemble fleet equipment with high fidelity, although it is functionally similar.

The trainer has the capability for simulating the simultaneous and independent movement of the following vehicles:

- 18 destroyers or submarines
- 16 aircraft (fixed-wing or helicopter)
- 1 aircraft carrier
- 9 drone anti-submarine helicopters
- 4 instructor-controlled target submarines

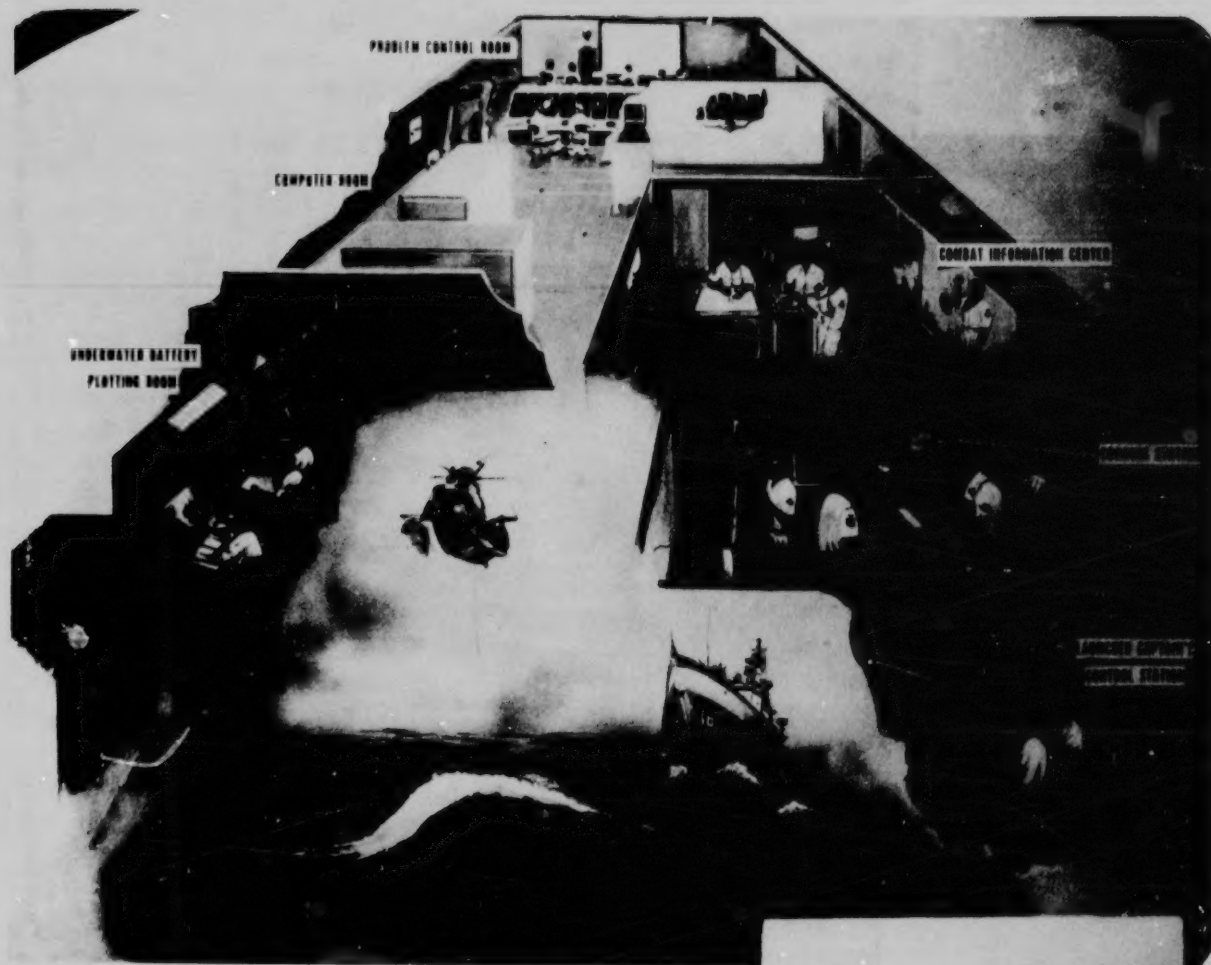


Fig. 2—Surface ship anti-submarine warfare (ASW) early attack weapons systems team trainer, Device 14A2

The critique room seats approximately 300 and serves as a pre- and post-mission briefing facility as well as an audience viewing area during problem evolutions.

The P-3C Operational Flight Trainer, Device 2F87(F), provides training for the aircraft pilot, co-pilot, and flight engineer across the range of dynamic flight profiles and malfunctions or in-flight emergencies which may occur. The pilot, co-pilot, and flight engineer stations duplicate the corresponding stations of the actual aircraft, including instrument panels, overhead panels, center pedestal, and throttle area. The trainer consists of the following major equipment items:

- Cockpit simulator
- 6DOF motion platform
- Model board for visual simulation
- Instructor station
- Computer and peripherals

Coordinated missions can be simulated using two cockpit simulators, or one cockpit simulator may be used in conjunction with the 2F87 Weapons System Trainer (WST) (not shown) for ASW team training. The WST contains the crew positions in the remainder of the P-3C aircraft which are principally involved in ASW tactics.

The final example represents a quite different approach to team training. The Combat Systems Trainer (CST) consists of van-mounted equipment used to simulate radar and electronic warfare sensors on board DLG16 and DLG26 Class ships. The CST provides training for radar operators, weapons control, electronic warfare, and combat information center personnel. The principal advantage of this approach to training is that participants occupy their normal duty stations on board ship, thereby eliminating any adaptation to some training-unique simulator.

SOME RESEARCH ISSUES

It should be clear from these examples that Navy tactical team training covers a broad spectrum in terms of levels of complexity. However, certain structural characteristics and behavioral requirements are common to most all teams of this type. Rather than attempt yet another definition of a team, I see much commonality among existing definitions. A team, in general, has the following characteristics:

- It is goal- or mission-oriented.
- It has a formal structure.
- It has members with assigned roles.
- It requires member interaction.

In comparing the attributes of a team to those of another type of group, Klaus and Glaser (1968) offer the following distinctions:

... a team is usually well organized, highly structured, and has relatively formal operating procedures—as exemplified by a baseball team, an aircraft crew, or ship control team. Teams generally:

- are relatively rigid in structure, organization, and communication networks,

- have well-defined positions or member assignments so that the participation in a given task by each individual can be anticipated to a given extent,
- depend on the cooperative or coordinated participation of several specialized individuals whose activities contain little overlap and who must each perform their task at least at some minimum level of proficiency,
- are often involved with equipment or tasks requiring perceptual-motor activities,
- can be given specific guidance on job performance based on a task analysis of the team's equipment, mission, or situation.

A small group on the other hand rarely is so formal or has well-defined, specialized tasks—as exemplified by a jury, a board of trustees, or a personnel evaluation board. As contrasted with a team, small groups generally:

- have an indefinite or loose structure, organization, and communication network,
- have assumed rather than designated positions or assignments so that each individual's contribution to the accomplishment of the task is largely dependent on his own personal characteristics,
- depend mainly on the quality of independent, individual contributions and can frequently function well even when all or several members are not contributing at all,
- are often involved with complex decision-making activities,
- cannot be given much specific guidance beforehand since the quality and quantity of participation by individual members is not known.

Rather than enter into a detailed discussion of some of the finer distinctions between a team and a group, I would like to focus on what I consider the most significant difference, i.e., the comparative structural formality of a team. Klaus and Glaser's baseball analogy seems reasonable. If all the professional shortstops and second basemen were placed in an urn, and one of each was randomly drawn, they would probably function as an effective double-play combination with a minimum of interpersonal adaptation. The key word is *professional*.

The implications for Navy team training are clear. I am firmly convinced that Navy team performance can be greatly improved by requiring higher levels of proficiency for individual subtasks.

However, the Navy training establishment cannot be content with arbitrary levels of team proficiency. We must produce *winning* teams. What distinguishes a winning team from one which is merely proficient? How do we observe it? Can it be measured? And perhaps more importantly, can it be trained?

Anyone who watched the World Series this year witnessed some very shrewd tactical decisionmaking on the part of the winning team's manager. A relatively mediocre pitching staff was carefully manipulated to challenge opposing hitters based on their comparative strengths and weaknesses.

Similar circumstances exist in most Navy tactical situations. That is, much of the variability in team performance is a function of individuals following established procedures. However, the team's success or failure is often attributable to the performance of the team leader, or decisionmaker. While it may be argued that those responsible for making key decisions (e.g., tactics selection) are also following procedures, I suggest that their tasks, in terms of human perception and information processing, often represent a much higher order of cognitive complexity. This poses numerous questions:

- Is it a structural problem?
- Is it a training problem?
- Is it a selection problem?
- Is automation the answer?

Another perplexing issue concerns the business of coordination. Training cadre have indicated that the main purpose of team training is to teach crew coordination. But what is coordination? Views of interaction and concepts of coordination vary depending on the referent situation and the assumptions made concerning crew behaviors. The literature tends to treat coordination from the standpoint of the tasks that must be performed within the team, whereas training personnel seem to view it more from the standpoint of the performer.

It seems that the nature of coordination is moderated by two classes of events. These are *established* situations and *emergent* situations. In an established situation, crew interaction occurs where roles are highly structured in terms of responsibilities and operating procedures. Coordination, in this sense, may be objectively identified as a form of procedure-following, as opposed to some unique behavior. Coordination also results when team members interactively perform in emergent situations where procedures and standards of performance are less clear. Here the activity of coordination has the appearance of more complex behavior. This behavior might be characterized by improvisation or extemporaneous response to emergent events.

Also of interest are the sociological dimensions of coordination which may affect performance. An established crew may, over time, exhibit some sense of "team awareness." This may take the form of anticipatory or supplementary adaptive behavior which evolves with increasing interpersonal familiarity.

While we may describe various dimensions of coordination, much is ambiguous. It appears that coordination in established situations is manageable and may be attacked much the same as other procedural activity. However, the inventive or sociological aspects of interactive behavior are more elusive and perhaps are worthy of further investigation.

While fundamental learning principles are implicit in any examination of team training, the topic of feedback has perhaps received the most attention. In reviewing the research literature on feedback, Kanarick, et al. (1972) concluded that "Performance feedback is unquestionably the single most important parameter in team or individual training." It seems reasonable that students need to know how well they are doing during training to insure that they are acquiring the proper information and that they appropriately redirect their efforts as error occurs. In practice, however, feedback is normally treated in a casual, nonsystematic way. How much, when, and what kind of information students receive seems to be a function of individual instructor practice rather than deliberate policy.

Unfortunately, many relevant questions which would logically arise in attempting to apply feedback to a team training situation are, at best, only partially answered by the research in this area. A partial list is suggested:

- When should feedback be provided?
- What information should be provided?
- Who should receive feedback?
- Should the amount of feedback vary according to the stage of learning?

CONSIDERATIONS FOR FUTURE RESEARCH

There are a number of factors which must be considered in formulating an approach for prospective team performance research: heuristic, economic, and political. Judging from the array of academic disciplines represented at this workshop, we will certainly discuss a number of technological areas which may be related to team performance. Our task, then, is to focus on those technologies which are the most promising for expanding our understanding of team behavior, and ultimately for improving team performance.

Given that finite resources will be available for team performance research, we must concentrate our efforts in a way which will provide the Navy with the greatest possible benefit. This will involve a scrupulous examination of existing team performance issues to prioritize our efforts in attacking those variables which may, in fact, be altered within the Navy system.

Political considerations cannot be ignored. It has been my experience that training managers and those tasked with the conduct of training are becoming increasingly suspicious of researchers. All too often, training establishments have hospitably opened their doors and bared their souls for those promising to improve their lot, only to be rewarded with confusing statistics and esoteric jargon. The results of our efforts must have credibility and utility in the eyes of the operational Navy, as well as methodological precision.

As members of the research and development community, we are challenged to come to grips with complex phenomena and improve the realm of team performance. I would not presume to specify the details of this effort, but I will offer a "jacks-or-better" skeleton which I hope will provoke thought and discussion.

Typically, research efforts begin with a hypothetical proposition which is then examined through systematic experimental design and analysis, leading to information, or at least inferences, concerning the topic of investigation. Efforts of this type have chipped away at team performance issues for some twenty-five years with mixed success. It appears that much useful information has been documented, yet little has been applied.

I recommend a pragmatic approach. I suggest the establishment of a team performance "laboratory" within the context of an existing Navy training facility. This could be done to service a new weapons system or to overhaul an existing training program. This model training facility should have two prominent features:

1. It should be designed based upon well established principles of educational technology.
2. It should facilitate comprehensive individual and team performance measurement, unobtrusive to training.

Some of the more salient steps in this approach are outlined as follows:

- Select, with Navy assistance, an optimum environment.
- Develop a model training system using fundamental principles of instructional systems development.
- Design comprehensive, automated scoring and recording capabilities for both individual and team performance.
- Gather data through the conduct of routine training.

- Analyze and account for that variability in team performance which may be attributed to individual performance.
- Formulate hypotheses concerning unknown sources of variance.
- Systematically manipulate the training environment to examine remaining sources of variance.

I suggest this field research approach for several reasons. First, I believe that team performance can be improved *today* by a judicious application of sound existing technologies. It seems that we expend a lot of resources for training research yet do very little to translate the outcome into something of tangible utility for the Navy.

Second, we will have established a valid, ongoing "laboratory" for continued examination of some of the more subtle or elusive determinants of team performance.

And finally, I believe that by bridging the gap between our more academic endeavors and the very real problems extant in team performance, we can do much to reinforce our credibility in the view of the operating Navy.

A LOOK AT TEAM TRAINING

Larry Nowell

I have been a member of many Navy teams, large and small, including Combat Information Center (CIC) teams such as Anti-Air Warfare (AAW), Electronic Warfare (EW), Anti-Submarine Warfare (ASW), Naval Gunfire Support (NGFS), Radar Navigation teams and Air Intercept Controller/Fighter teams.

I have also spent eight years in schools teaching operators and teams, I have helped to develop consoles to be used for basic operator training that are designed with training in mind, and I have worked as a subject matter expert, providing input to system design and writing courseware for the Prototype Air Controller Training System (PACTS). Thus, it would be possible for me to discuss a whole range of team training issues. However, in order to establish some concrete ground for discussion, this paper will be restricted in scope to concentrate on the AAW team in the CIC aboard ship.

We can begin by visualizing a lone ship sitting out in the middle of the ocean, a fighter stationed overhead. Inside the ship is a large room designed to display the tactical situation. This room, the CIC, serves as the tactical nerve center of the ship. The teams in the CIC gather, display, evaluate and disseminate information from radars, identification equipment, electronic warfare equipment, communication nets, operation orders, and the like. Within the CIC, a detection is made by the AAW team of an unidentified aircraft closing in on the ship. If the contact cannot be positively identified as friendly, the command and control team may assign the air controller/fighter team to intercept, with surface-to-air missile teams alerted as backup.

This description of one ship, one fighter, one air controller, and one target is oversimplified; nevertheless, it gives some indication of the complexity of multiple-team activities within the overall CIC team. Teams in the Navy range in size from two-man teams, through the combination of all the teams of a ship, and upward through the coordination of task forces. The intricacies of team training for the CIC of one ship alone become more obvious if the example ship's AAW team must engage additional hostile aircraft concurrent with the surface action team detecting closing high-speed surface contacts and the ASW team gaining SONAR contact on a submarine. Placing the ship within a multiship task force emphasizes the enormous complexity of team training to prepare for confronting an actual multi-threat environment.

Prior to receiving any team training, the operators/evaluators receive individual training to prepare them for their watch stations aboard ship. Officers receive training in the areas of tactics and evaluation, while enlisted men receive training for the operator and operator supervisory positions in the CIC. Enlisted training starts with basic training in navigation, plotting, and console operations and continues into the more sophisticated areas of anti-submarine air controllers and air intercept controllers. Additionally, on-the-job training aboard ship helps prepare the individual to be a team member. While in port, some ships may have the services of vans that can be used for pierside team training. An example of a new

pierside trainer is the Multiple Link Test and Operational Training System (MULTOTS) which provides (1) environment simulations, (2) participating unit simulation, (3) test/exercise monitoring, and (4) equipment performance checking.

Depending on the ship's operating schedule, the crews may also have extensive team training at the Fleet Combat Training Center (FCTCP) in San Diego. When possible, ships send their teams for training each quarter. The training lasts five days and consists of lecture presentations and exercises that are run in a mockup that is laid out similarly to the CICs aboard ship.

The equipment used to train the teams is the Tactical Advanced Combat Direction and Electronic Warfare (TACDEW) training facility. Instructor/staff that conduct and support the team training include:

Mockups

- Two evaluators for each mockup

Program Control and Evaluation

- One overall coordinator
- One program developer for the air picture
- One program developer for the surface picture
- Two electronic warfare instructors
- Four to five pseudo-pilots

The team training course begins with lectures on the expected enemy threat evaluation, Combat Air Patrol (CAP)/missile coordination, contact/raid reporting procedures, etc. Special lectures may be arranged as requested by the ship, but the majority of the time is spent in the mockups running simulated battles, obtaining available data, detecting targets, assigning fighters and missiles, operating equipment to the highest degree attainable, and working on making the right decisions. The difficulty of the training problems is adjusted according to the individual expertise of each team.

The team's evaluation is done by the FCTCP staff grading the individual operator positions. The first couple of days are spent getting the teams accustomed to the school's consoles; their presentations and controls may be different from the ship's. Informal evaluation and some individual instructions are also given while the exercises are in progress. On the fifth day of training, formal evaluation of the team is based on a battle exercise in the mockup. Individual operator positions are monitored for performance of the station by over-the-shoulder observation. Final team grades are the sum of the grades assigned the individual positions plus or minus a "fudge factor" applying the observer's "gut feel" for the team's performance.

The description above provides some indication of the composition of a typical team and how it might be trained. I will use this example to cite some limitations I perceive in the present system.

The mockups being used at the school are designed to represent different typical CIC layouts. The consoles the operators use may be different from the ones they have aboard ship. There are three consoles in use in the fleet today, and each has its own unique features. Thus, operators being taught basic console operation at NTDS school are taught a specific console, while their ship may have another type, and the team training school may have a third.

The computer programs used to run the consoles are developed by the Fleet Combat Direction Systems Support Activities on the east and west coasts. The programs for smaller ships are written on the east coast; those for large ships, on the west coast. Ships stationed in San Diego or Norfolk would have a mix of each of the programs. Although the programs are written to be compatible between consoles and ships, the displays, presentations, and controls for individual operator stations differ.

Functions on the operator panel are not standardized. These functions should be performed by the operators without looking at the buttons or labels. Moving an operator from one console to another might be compared to presenting a secretary with a new typewriter—with a different keyboard.

Team evaluation seems to have several weaknesses:

1. Team evaluations are subjective. Some instructors may never have had any AAW team experience, yet they may be alternating in evaluating with instructors who have had extensive AAW team experiences.
2. Some operator positions are graded only part of the time. There is never enough staff to grade all positions all of the time; only periodic sampling can be done.
3. There is no preplanned script that allows for structured, controlled progress with established performance standards. Three staff members vary the complexities on-line, while the battle problem is developing. Generally they give the team as much as they can handle, then a little bit more, then they back off.
4. Does individual performance reflect a team's efficiency? The criteria that define a team's performance may never have been adequately identified. Individual scores may not truly reflect team performance.
5. Some team components necessarily may be missing. The ship's weapon teams are not trained with the CIC teams at the training center. The ship's own weapon capability is not necessarily reflected in the simulated weapon capability in the mockup. Teams may train with unrealistic weapon capabilities, and the interface between teams is lost.

The pseudo-pilots who fly the simulated aircraft are team training staff; they have no pilot training, nor do they have a feel for the correct pilot/aircraft reaction for a given situation. They tend to react with indifference to the simulated situations. This humdrum attitude hinders development of the vital emotional factor of a good air controller.

The initial basic training, A school and NTDS school for the individual operators before shipboard duty or team training, does not take into consideration the other operator/evaluator positions the trainee has to interact with. The operators are taught about the equipment and how it works; the actual job is left to be learned aboard ship. The cognitive processes the operators perform are grossly neglected. As an example, the air controller is trained without considering interactions of the other air controllers, the air controller supervisor, or the Ship's Weapons Coordinator (SWC).

The smaller teams that make up large teams have no controlled training. This intermediate training is overlooked. For example, the air controller, other controllers, AIC Supervisor, and SWC team have no facilities for training together as a

team. Smaller teams need training before being thrown into total team training. The ship teams being trained have no simulated model of other ships with which to learn the proper interactions prior to working in a multiship environment. There is no systematic method to ensure the team can properly respond to individual threats before being put in a multithreat environment.

In the face of these perceived team training deficiencies, it is important to take into consideration some of the positive advances in providing more adequate training. The Office of Naval Research (ONR) and the Naval Equipment Training Center (NETC) are issuing research and development contracts to identify, research, and develop new training approaches and technologies to help provide training solutions.

One contract that seems particularly applicable to these sessions is titled *Prototype Air Controller Training System, Air Intercept Controller*, or PACTS AIC. This contract is designed to develop an experimental training system for the AIC, a CIC team member. The experimental, standalone training system is quite sophisticated and includes such diverse features as:

- Automated speech
 - Computer-generated speech
 - digitized speech
 - synthesized speech
 - Computer speech understanding
 - continuous speech recognition
 - semantic logic
 - contextual logic
- Automated instructor
 - Automated performance measurement
 - Adaptive syllabus control
 - Diagnosis
 - Prescription
 - Remediation
 - Student record keeping
- Simulated work station
 - Simulated UYA4/V10 NTDS console
 - Simulated tactical environment
 - Simulated aircrew training environment
- Videodisc-based audiovisual system

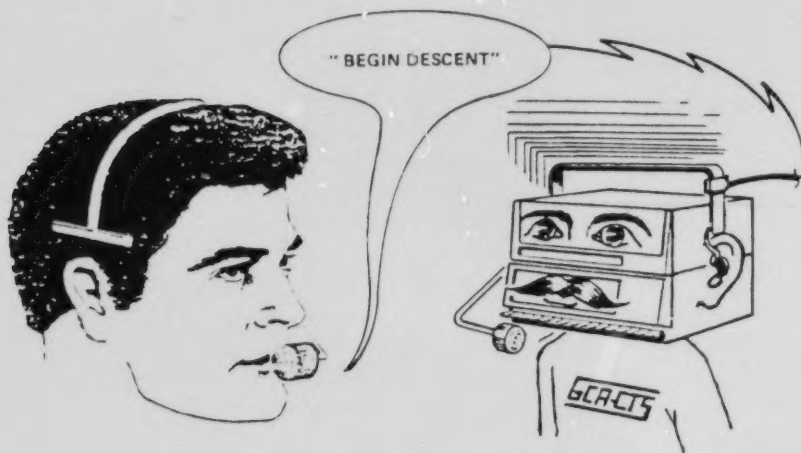
Below we take a brief, but more in-depth look at some of these features.

Automated Speech

Automated speech includes all aspects of computer generated speech and human speech recognition. This is one of the principal facets of PACTS AIC (see Fig. 1).

Speech will be generated using both playback of digitized human speech which has been stored on computer disk and artificially generated (or synthesized) speech produced using a Votrax speech synthesizer. Speech understanding will be based on state-of-the-art continuous speech recognition hardware from Nippon Electron-

YOU SEE, THE COMPUTER LISTENS WHILE YOU ARE SPEAKING



THEN IT COMPARES WHAT IT HEARD WITH ALL THE PATTERNS OF YOUR VOICE THAT IT HAS ON FILE

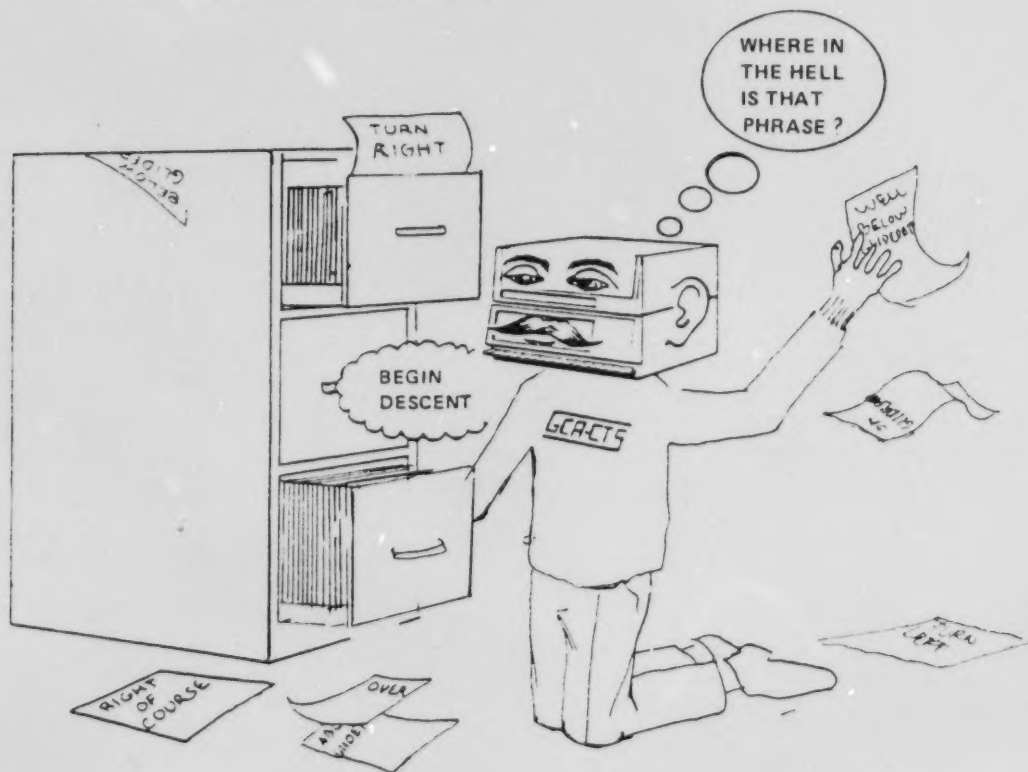


Fig. 1—An informal view of automated speech recognition

ics Corporation (DP-100) and a speech understanding program designed to assist the recognition process. The speech understanding works through the application of contextual, syntactical, and semantic decisions to limit the recognition possibilities.

Speech recognition is being used to replace the pseudo-pilot and provide inputs to performance measurements.

Automated Instructor

PACTS AIC is designed to be a standalone system. This means the curriculum must be presented, the student's learning measured and scored, and the syllabus adaptively changed to provide an optimum learning experience (see Fig. 2).

The PACTS AIC curriculum is designed around contemporary learning, reinforcement, and motivation principles and theories. The curriculum is being developed using principles of instructional system development (ISD) by a team of training analysts, courseware specialists, and subject matter experts.

The system, as currently envisioned, will provide the learner with a three-week instructional sequence which changes according to his problems, successes, and identified needs.

The AIC trainer has several features which can prove advantageous to improving team training. Primary among these are the carefully defined instructional objectives, conditions, and standards; the precise automated measurement and scoring of learner performance; the computer-based adaptive curriculum; and the simulated performance environment including work station and personnel models.

Through the application of current instructional system design principles it has been possible to define the objective of the AIC's job in conjunction with the CAP. Once this is defined, it becomes immensely easier to identify what the AIC's duties are and, more importantly, when the AIC is doing his job well. Instructionally, the job can be further subdivided by the identification of the behaviors, conditions, and standards which make up the total job.

The automated measurement and scoring, part of a computer-based automated instructor, also is based on these identified behavioral criteria. Instructor grading, as noted earlier, has been fairly subjective. The human instructor does not have the ability to attend to every facet of learner performance and usually does not have an adequately identified set of criteria from which to make judgments. The computer in the PACTS AIC has the criteria, patience, and attention to detail that allows in-depth measurement and analysis of learner performance. This provides an objectivity not available in most instructional environments, an important advantage for training.

The precise scoring of learning behavior allows the application of complicated learning achievement and retention algorithms. These algorithms identify the sequence of instruction and practice alternatives available within the curriculum which will best meet the learner's needs, based on previous performance. The automated instructor uses diagnostic-prescriptive-remediative sets to make decisions about learner advancement, remediation, or need for more work on the present topic.

All of the instructional system works in conjunction with a medium-fidelity simulation of the AIC's performance environment. This environment includes a simulated console, simulated displays of aircraft and missiles, and, perhaps more

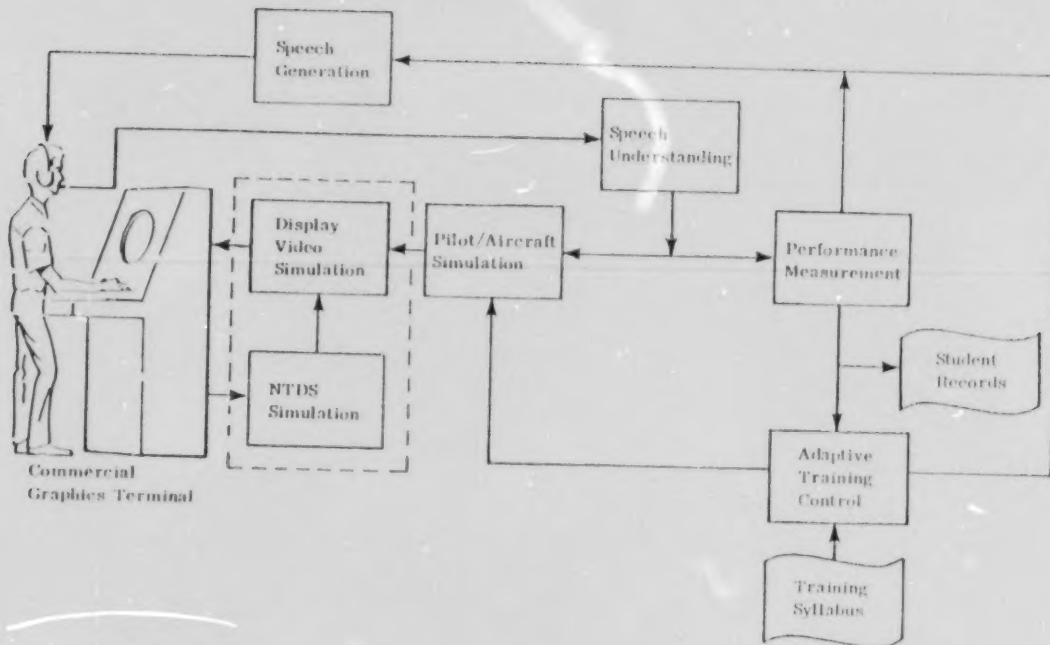


Fig. 2—Design components of PACTS AIC training system

important for team training, computer models of other team personnel including the aircrew, tracker, height-size operator, and Ship Weapons Coordinator (SWC). Use of these models provides a highly interactive, good-fidelity training environment without having to involve "canned" scenarios based on audio or video tape or the use of technically non-trained personnel acting as "pseudo" team personnel.

Considering both the perceived deficiencies and the work that is currently in progress, there are still some additional areas I would address as additional problems or possible alternatives:

1. Individual operator training is a must before any effective team training can take place. Incidentally, Meister (1976) has indicated that as a general rule for teams, individual training should come before "team-specific training except when specific procedures for coordination must be taught." The basic skills for each operator position have never been adequately identified, recorded, and incorporated into a training program. Operator schools, basically, are still teaching equipment. One must provide for this basic operator training, or it may never be done. The cognitive processes need to be examined. For example, one of our training analysts suggests that operators may be using trial-and-error processes on some console button actions.

2. The consoles and programs in use must be standardized or a training center that will allow for individual differences must be designed. Standardization is, of course, one of the most basic concepts behind the military standards for human factors engineering. An attractive alternative may be to train aboard their own ships with a MULTOTA-type van.

3. The standards for evaluating team performance must be defined or the summary of the individual operator scores must be confirmed as a valid measure. A similar recommendation was made in a review of team training by Collins (1977). Performance measurement should be automated to ensure objective grading while reducing human observer/ evaluator requirements.

4. Models should be developed to simulate other operators, other teams, and other ships and to integrate the man into his immediate team, as well as provide the interface with other teams in the CIC and with other ships.

SIMULATION-BASED TEAM TRAINING

H. McIlvaine Parsons

In this paper, I could address myself to complex experimentation as a method of investigating team performance. However, because I once wrote a book covering that topic (Parsons, 1972), there seems little point in repeating what those at Rand presumably have already examined. I would, however, like to mention two points. Both concern a potential research facility. In my book, I urged that the horse precede the cart and the research program be created before designing the facility, since in those instances where the facility was built before the owners knew what they would do with it, no significant research resulted. This happened at least twice where the same organization was funded not only to build and operate the facility but also to conduct the research in it. (The outcome might differ if the research program consisted of ongoing components in which a range of organizations could engage through competitive or sole source proposals, though such an arrangement would entail other complications.)

My other point regards the scope and flexibility of such a facility. A dedicated laboratory to investigate a particular system must incorporate the simulated sensors for, or inputs into, that system, the activities in that system, and its outputs and simulated effectors. For many kinds of systems (and teams), particular hardware will be required, limiting the use of the laboratory to investigate other systems. A general-purpose laboratory seems restricted to investigating systems using a general-purpose computer (or one that such a computer could simulate) and alphanumerics (or graphics) for inputs/sensors and outputs/effectors. The kind of system to be represented in a laboratory determines the kind of laboratory facility to be created. Except for deliberative and planning teams, team performance is always embedded in some man-machine system.

Instead of probing further into such matters, I would like to take an empirical approach to team performance by discussing past experience that has lessons for the future. On the chance that few readers have actually taken part in simulation-based training, I will begin by describing a training program I once developed and installed with some colleagues at the System Development Corporation (Parsons, 1960). Then I will review some of the problems that characterized SDC's System Training Program (STP) for the Air Force as I analyzed them in a report to the Navy (Parsons, 1964).

Between May 1959 and May 1960, the SAGE ECCM training program was developed, tested, and installed in 10 air defense direction centers and 15 long-range radar sites. The principal vehicle for training was exercises in which attacking bombers and their electronic countermeasures (ECM) were simulated on the operational displays at the radar sites and direction centers. Teams of operators at these locations were trained to apply counter-countermeasures so the attacking aircraft could be detected and tracked. More than 2000 officers and airmen were trained in 100 exercises as well as 135 briefings. Subsequently the program was extended to additional direction centers and radars and trained still more individuals.

SAGE (an acronym for Semi-Automatic Ground Environment) is or was the system of nationwide (and Canadian) air defense based on digital computers as well as people, replacing a system based on analog computers, which in turn had replaced a wholly "manual" system. In a complex network, positional data about the radar echoes of aircraft were sent from a number of radars to the computer at the direction center, where surveillance operators at displays and the computer jointly established the aircraft tracks. Hostile aircraft could jam the radars so that the aircraft's direction but not its range could be ascertained, or they could drop chaff (aluminum foil) that produced false echoes.

Although the radars had ECCM devices, except for some residue from STP for manual air defense, SAGE had no training to operate them, nor had the direction center personnel received training to apply ECCM. A multibillion-dollar system was operational but virtually helpless in the event of a Soviet air attack, which would certainly employ powerful and sophisticated ECM. Why this vast command and control system was designed, built, and installed in this manner is not the story I intend to tell, but when we started to develop our ECCM training program, little was understood about what ECM might do to SAGE. One of our tasks was to find out, so we could know what to train for.

Let me make it clear that we were able to complete our job in twelve months only because there was already an overall system training program functioning in SAGE, with concepts taken over from manual STP, non-ECM simulation in place at both the direction centers and some of the radars, and SDC representatives helping the Air Force conduct exercises at the direction centers. These exercises had not included the radars or the personnel there, however. Due perhaps to some computer mystique, no one paid much attention to the system's front end, its sensors.

The SAGE ECCM Training Program yielded a "model," somewhat idealized, that had not been previously explicated. It has ten steps:

1. Training requirements.
2. Training methods
3. Exercise configurations.
4. Simulation equipment.
5. Training materials.
6. Exercise content.
7. Temporal aspects of exercise.
8. Exercise-aiding techniques.
9. Program installation.
10. Post-installation development and support.

TRAINING REQUIREMENTS

The first question was, of course, whether there should be a training program for ECCM. Here there was no argument. The Air Force was worried about its helpless giant. Alternatives to training, including team training, are automation and human engineering design. Expectations that the SAGE computer, or rather its programming, could handle the ECM problem turned out to be just that, expecta-

tions. Only modest help could come from human engineering, even if this human factors function had received its due in SAGE.

The next question was who should be trained. Opinions were strong but hardly unanimous, and they were held in the absence of empirical data about ECM effects on the system. There was some surprise when we showed that more than 43 percent of the personnel who would be operationally active in an ECM environment in the event of the attack were at the radar sites. It was also easy enough to show that at the direction centers the key set of operations was surveillance, not the weapons control that had received most attention in SAGE STP.

A team of SDC investigators went to a non-operational SAGE direction center and two of its radars. For nine weeks, with about 150 hours of test time, we put simulated targets and ECM through the radars into the SAGE computer and its displays. Operational personnel at all three locations served as our test subjects. We were able to demonstrate that the radars were crucial ECCM elements, and we performed task analyses that indicated which tasks were individual and which were interactive; the latter especially needed team-type training. Some of these involved the formulation and transmission of commands and information, others the allocation and sharing of workloads. We rated individual and interactive tasks according to four criteria: (1) difficulty; (2) criticality for system mission performance; (3) amount of pertinent previous interactive or individual training on-site personnel would have received prior to the start of the program; and (4) amount of pertinent training the personnel would get from normal duties or other sources if the program were not introduced. The determination of training requirements—the first step for any team training program but at the time a seemingly novel one—was furnished to the Air Force and SDC management and upon approval provided the rationale for ECCM training in SAGE.

TRAINING METHODS

With little debate, simulation-based exercising was adopted as the primary training method. The manual, AN/GPA-37, and SAGE System Training Programs had been based on synthetic exercises as a result of the Rand System Research Laboratory experiments from which STP had sprung. An alternative would have been to have exercises with actual aircraft—the interceptors themselves that would have to repel an attack, and SAC bombers, representing Soviet aircraft jamming and dropping chaff as the Soviets might do. (Live aircraft ultimately were used to test SAGE reactions to ECM, and the highly classified results became one of the best-kept secrets in the military history of this country.) In addition to avoiding security problems that any "live" exercise might incur and eliminating the chance of disrupting air and ground communications in the exercise area, simulation could generate realistically large numbers of bombers and realistic amounts and varieties of ECM with a greater frequency and diversity and with much less expense than would be possible with actual aircraft.

Dependence on simulation in non-ECCM (as well as ECCM) exercises meant that interceptor pilots did not take part in an exercise, though it would have been possible to scramble and direct interceptors against synthetic targets with which the pilots and airborne fire-control system would never make contact. By defining

the system as excluding its effectors (the interceptor aircraft), STP planners were able to omit interceptor pilots from exercises and still cling to the principle of "exercise the system as a whole." The exclusion of interceptor aircraft as well as airborne simulated bombers from ECCM exercises meant that interceptor pilots failed to receive training in trying to counteract jamming of ground-air communications and of airborne fire control systems. In retrospect, we believe live exercising should have supplemented synthetic exercising.

Because electronic countermeasures were relatively unfamiliar to many SAGE personnel, we decided to include in the training program primer-type reports and extensive briefings and lectures, especially during the installation.

EXERCISE CONFIGURATIONS

In non-ECCM SAGE STP exercises, configurations varied between single direction centers, groups of direction centers netted with a combat center, and sets of combat centers and their associated direction centers up to all those in the United States and Canada. But as already noted, the system's effectors (aircraft) and sensors (radars) did not take part. Despite some raised eyebrows, we decided to concentrate on one subsystem at the direction center—surveillance—rather than the entire center, and to include the radars as another subsystem. In ensuing years, my branch at SDC developed non-ECCM team training separately for weapons and surveillance operations in a direction center. The WEST program, for weapons team training, was field tested in a SAGE sector and adopted by the Air Defense Command. Its advantage was that the synthetic inputs to intercept directors and their supervisors could be systematically planned and feedback could be given to operators. Otherwise, inputs depended on prior processing in the surveillance subsystem.

SIMULATION EQUIPMENT

As everyone knows, in training programs the simulation equipment is too often designed and produced first and the training program second, so the program must be tailored to the equipment. In the original development of STP for the manual air defense system, however, the proper order of events was followed. The AN/GP-T2, designed for the program, consisted of an array of flying spot scanners and the electronic target generators. The scanners picked up signals from marks on 70mm film, on which each frame represented one antenna rotation. The marks, indicating azimuth and range of targets (and chaff), were placed on the film by means of a computerized production system. To improve ECM simulation, an Anticountermeasures Trainer (ACTER) was added to give four types of simulated jamming. We used this set of equipment to simulate targets and ECM through the radars and thereby at the direction centers, after making some modifications and testing the equipment for SAGE exercises. Meanwhile, entirely new simulation equipment for ECM/target inputs was proposed to the Air Force, though to my knowledge it was never funded. All of this simulation differed from what was being used for exercising SAGE in a non-ECM environment. For that purpose, target positions were stored on a magnetic tape and thereby entered directly into the AN/FSQ-7 comput-

er at the direction center. If this technique could be modified to show the effects of ECM at direction center consoles, the surveillance section could be exercised without inputs from the radars (and without involvement of and interactions with the radar personnel). Eventually, methods for doing this were developed.

This discussion of simulation techniques may seem a little complicated, but I have purposely included the details to indicate that simulation itself can be a complex matter. It is necessary to consider not only cost and feasibility but also verisimilitude, that is, fidelity modified by the trainee's reactions. How great should the verisimilitude be? To generate interactions within a team it may have to be less than to provide practice in individual discriminations. Let us not forget that in the Rand SRL experiments, operator displays consisted of IBM 407 printer multifold paper with the digit 1 representing targets, rather than real Plan Position Indicator cathode ray tubes with radar-like "blips." For ECCM training, considerable verisimilitude was required because operators had to "read through" noise to discriminate targets. But the verisimilitude problem was complicated further by having to figure out what actual Soviet ECM would look like on radar and SAGE displays. The training developers had to know quite a lot about radars, ECCM "fixes" (anti-jamming devices) associated with radars, equipment converting radar signals into digitized data going from radar to SAGE computer, and operations of the computer and its associated displays. This material was too puzzling for many of the clinical, social, and cognitive psychologists at SDC and called for closely working with engineers, yet the same kinds of problems may arise in other team training development.

One of the special equipment problems in depending on inputs from radars for SAGE ECCM training was maintaining live air defense while conducting synthetic training exercises. Operational commanders did figure out solutions when they came face to face with the necessity.

TRAINING MATERIALS

Production of simulated inputs for exercises constituted another technical complexity. Fortunately for the ECCM program, SDC already had a sophisticated production system to create and distribute such inputs. Dozens of aircraft tracks for an exercise were first formulated in general terms and then drawn from a library or were manually generated and combined with such factors as timing, spacing, jamming, and chaff, to present a realistic air picture. For the final output, the production computer also received environmental data that affected track characteristics—radar and other air defense aspects, weather patterns, flight plans, aircraft characteristics. Track manuscripts were translated into cards for an IBM 709 computer that then produced a magnetic tape and various printed materials with EAM and EDPM equipment. Special equipment converted the tape contents to the 70mm film for the AN/GPS-T2 and ACTER. Some exercise aids, such as scripts and messages, were prepared by hand. Each radar in a SAGE sector received a different film, but radar echoes of aircraft detected by two or more radars at the same time had to have identical geographical positions on different films. (Similarly, magnetic tape inputs going directly into different SAGE computers had to have synchronized positional signals. Synthetic inputs of this type totaled more than 300 different exercise hours per year.)

For installing the ECCM training program, we used two "problem" films developed for our test phase. One was a demonstration of various ECM and aircraft parameters, the other a tactical exercise in a mythical but realistic environment. We also produced the primers already mentioned and slides for briefings. The mythical environment made it possible to use this input at different locations, since it did not have to be tailored to a particular area as did all regular exercise inputs.

EXERCISE CONTENT

For subsequent ECCM exercises, as in non-ECCM exercising, it was necessary to consider three criteria, not necessarily coherent, in designing their content. One was realism, to (1) maximize positive transfer of training from exercise to wartime situation, or minimize any negative transfer, and (2) get the trainees and their supervisors to accept the training, that is, to refrain from resisting it on grounds that it was not realistic. (However, we had no problem along this line with our mythical environment during installations, when we made it clear that we were not trying to represent the particular geographical area.) The second criterion was to present input loads or situations requiring individual and interactive performance seldom otherwise activated. This criterion could conflict with realism, especially in large-scale exercises, since in warfare some locations could be expected to encounter heavy input loads, others light ones or none at all. The third criterion was difficulty level. Should increasing levels of difficulty, requiring higher degrees of skill, be introduced as training progressed? Might some other pattern be more effective? All team training must face this uncertainty.

TEMPORAL ASPECTS OF EXERCISE

How long should an exercise continue? How often should exercises be held? Should the schedule be regular or irregular? Is any particular time of day or day of week preferable? We found few ground rules as answers to these questions, yet they are very real concerns for developers of team training, provided those developers carry their concerns over into the operations of the program they have created. Certain durations may reflect better than others a "real-life" situation. Frequency may depend on the competition from other duties and activities or, in a computerized system, the availability of computers. For ECCM training, we recommended direction-center/radar exercises lasting between 60 and 90 minutes, based on computer availability and the probability of exercise aborts because of equipment failure. At least four exercises per month for each direction center and radar would exercise each crew at each location at least once a month.

EXERCISE-AIDING TECHNIQUES

From its inception, the STP had emphasized (1) observing and recording individual and team performance in various ways through a Training Operations Report (TOR) and (2) post-exercise crew discussions ("debriefings") to get feedback

and formulate improved procedures as part of "problem-solving." We elaborated on and supplemented these for ECCM training and at the same time abandoned the TOR and debriefing as unitary concepts. We had to give more attention to individual ECCM skills because they were in short supply. Formulation of operating procedures could not depend entirely on debriefing discussion because operational personnel knew so little about ECM effects on their system. TOR (Air Force) observers had difficulty in monitoring performance because they too were unfamiliar with ECM and ECCM. Further, unlike non-ECCM SAGE exercises, it was not possible to associate through identification numbers the aircraft tracks on the displays with those in the planned inputs in order to determine surveillance success; this is relatively easy to do when inputs go right into the computer.

However, TOR observers at both the direction center and the radars were given maps and lists showing aircraft and ECM (jamming or chaff) inputs at 5-minute intervals—a standard technique in manual STP—so they could preview the exercise and try to match outcomes with inputs. In addition to placing observers at the radars as well as the direction center, we had others monitor the communications between these locations. Operators self-recorded their performance on forms given them. TOR observers and supervisory personnel kept logs of inputs, actions, and outcomes. Some innovations had particular interest: (1) The ECM (ACTER) inputs were occasionally turned off so operators could see track inputs unaffected by noise and evaluate their ECCM actions. (2) Comments on developments were made over the loudspeaker at the direction center to explain what was occurring at the radars, describe delays in target detection or reveal non-detections, and discuss procedural variations, attempted or potential. (3) Observers coached individual operators by pointing out effects of their actions, rather than waiting for the debriefing. (4) Displays were photographed during the exercises and the pictures projected afterward to show the effects of ECM "noise," something that is very difficult to do in words. (5) Debriefings included one on the telephone net between the direction center and the radars with participants at their duty stations, as well as separate debriefings at the radars, in surveillance sections at the direction center, and among supervisors. To the extent thought appropriate, TOR observers compared box scores of track initiation times, numbers of targets tracked, and durations of tracks with input information to derive detection lags and overall tracking effectiveness. One question concerns how much box scores (like the score in a football game) can aid training. Box scores are feedback, but feedback not correlated with particular individual or interactive performance—though it is useful for evaluating the progress of training and those system aspects that especially need it, and perhaps also for motivating trainees.

PROGRAM INSTALLATION

The best team training program ever designed and packaged may never have its proper effect unless the gap is bridged between the place and organization where it is developed and the field locations and agencies where it is supposed to be used. Who does the bridging? Often this is a no-person's land, outside the reward structures of either field or home office. Field people, civilian or military, salesmen or local representatives, may resist change, especially if substantial effort is required without overt indications of immediate gain—as is the case in training. Technical

continuity is needed as well as field acceptance. Fortunately, SDC had a functioning field organization, with one staff member familiar with both ECCM and field operations who joined forces with us and visited all the prospective locations where our installation team, with supporting engineers, initiated the program.

The first step had been a SAGE ECM Training Conference in Santa Monica, attended by a variety of Air Defense Command officers whom the program would affect. The conference reviewed all the plans we had prepared for program installation and subsequent training. We assumed responsibility for bridging the gap between development and operations.

The bane of simulation-based team training is the uncertainty that simulation equipment will function as it should. One cannot overemphasize how important it is to make certain this equipment will be properly operated and maintained not only at the outset but as long as the program continues. SDC engineers were key individuals among the installers, giving instruction in alignment and maintenance to military personnel at the radars.

After the first installation, which took two weeks, each installation required almost a week. The installers traveled on the weekends—from New York, New Jersey, Maine, and Virginia, to Michigan, Iowa, Minnesota, Nebraska, and South Dakota, and then on to Washington/Oregon. SDC field units at each direction center took part. There were both demonstration and tactical exercises, as already noted; in the former, the SDC exercise director maintained operational control and conducted a running commentary over the loudspeaker system at the direction center and the telephone lines to the radars. As mentioned above, there were 100 exercises and 135 briefings at 10 centers and 15 radars.

POST-INSTALLATION DEVELOPMENT AND SUPPORT

Development responsibility does not terminate with installation of the product. Although the field operations people at SDC took over the remaining installations and continuing operations on schedule, together with the Air Force personnel who actually conducted all STP exercises, further development took place. This took the form of writing a training handbook, formulating additional training requirements and methods to incorporate new SAGE functions and personnel, recommending exercise input features, and establishing a training course to train the trainers and help them feed back their training experience.

PROBLEMS

Odd as it may seem, the only time, to my knowledge, that the STP underwent a comprehensive analysis was when the Office of Naval Research asked me to write about what the Navy's Anti-Air Warfare Training Program could learn from Air Force system training experience. A cynic might say SDC had a good thing going that did not need close examination or quality control. However, such analysis might have helped extend system training to systems other than air defense, something SDC tried to do without much success.

TRAINING OBJECTIVES

No team training program can be simply transferred from one kind of system to another, on the assumption it is needed. As in system conceptualization and acquisition generally, it is necessary to look first at training objectives (the mission), then training needs (requirements), and then implementation (methods and media, constraints, etc.). Obviously the objective of team or system training is to improve performance, but when such training consists primarily or largely of conducting exercises, there are two subobjectives: (1) raising skill levels through practice, and (2) procedurization, the development of procedures. Procedures are alternative sequences or aggregates of actions, in which skill level can vary. Procedurization is a little-realized outcome of team exercising in a training program, though it can be an important one. It was highlighted in the STP but at times was confused with skill enhancement. The two objectives should be distinguished during the training development process because they differentially affect the methods and media to be used, the techniques to enhance the products of exercising, and the inputs to be selected for exercises. Training specialists must also bear in mind that exercises are not the only way of selecting among procedures, or creating new ones—just as in general problem-solving may be done by means other than direct experience. Since procedurization can result in wrong as well as right procedures, there should be some systematic method of testing procedures that originate in training exercises. There should also be methods of getting new procedures accepted by higher authority and of having them disseminated to other units encountering similar operational problems. In the STP there was very little documentation of procedures arising from exercises, despite the emphasis STP placed on this outcome.

My description of the ECCM training program for SAGE illustrates the value of carefully diagnosing training needs rather than relying on offhand judgments. Task analysis should examine not only interactional and transactional activities among personnel but also the flow of information into the system and its transformations in proceeding from one subsystem to another. Such analyses will indicate the complexity, variability, and predictability of inputs to personnel in their operational tasks, both individual and interactive. Team training must consider interactions between teams as well as between individuals (linear or parallel in nature): Team interactions depend on subsystem interactions. Team performance, especially at the front end of the system, and the consequent need for training, will depend on what the opponent is doing, if there is an "anti-system"—as there usually is for military teams. However, if the military organization happens to suffer from an "ostrich syndrome," which can happen, the anti-system may figure too modestly in team training.

TRAINING METHODS

When a team training program is based primarily on exercises, one should not neglect what can also be accomplished through instruction. Knowledge about the system can be gained through coursework in system operations, in computer program capabilities and limitations, and in relationships with other systems and the environment. However, the effectiveness of instruction is progressively limited

with increasing difficulty in predicting and verbalizing interactional procedures and information transformations.

As indicated in the ECCM story, "live" exercises represent an alternative to simulation-based exercises, and there are various possible combinations of the two. In STP, SAGE operators worked at their operational consoles, not at mockups. STP relied primarily on the simulation of radar echoes, but it also used scripted messages for voice or paper communications. Training personnel at the other end of a phone represented other agencies, such as an interceptor base or an adjoining SAGE sector. By and large, SAGE system training simulation aimed at identity—the simulation was as similar as possible to the matter represented. But there are other types of simulation that can be exploited in team training: (1) replica (scale modified); (2) replica (abstracted); (3) pictorial or graphical; (4) analog (functional); (5) verbal (words, numbers); and (6) computer bits (see Parsons, 1978). These can be combined in many different ways and have been in experiments, but team training developers, at least in the SAGE era, have failed to examine their potentials. Much analysis and inquiry should go into different methods of simulation and their possible combinations for team training, with emphasis on transfer of training related to interactional skills, information transformations, and procedurization.

EXERCISE CONFIGURATIONS

For a large system, team training exercises can have a variety of configurations, and possibilities should not be obscured by loose terms such as "functional whole" or "system" training. One should ask what advantages or disadvantages accrue to one particular scope or configuration rather than another. Expanding the scope of a training exercise increases the number and variety of actual information and control inputs and their transformations—more people can receive training for executing or procedurizing more actual interactions and transactions and responding to more actual input transformations. The alternative is to have simulator personnel interact with participants and to preprocess and simulate transformations of input information. How well this is done depends on the role-playing skills of the simulator personnel and the preprocessing. Simulation of input transformations is appropriate depending on how well such transformations can be predicted (including those that occur in dynamic interchange) and then introduced as inputs to the exercise. This can be surprisingly difficult. Procedurization probably benefits more than skill improvement from the inclusion of actual interactions and transformations in an exercise.

However, useful training can come from limited configurations—even a single person with extensive supporting simulation, if there is a need for that individual to have significant interactional skills and participate widely in procedurization. These individuals have tasks such as distributing the load, establishing priorities, selecting among alternative facilities, making complex or far-reaching decisions, and collecting information from diverse sources. Team training can emphasize this "nodal position" configuration. Another configuration is the subsystem, but subsystem exercises are not feasible with live inputs; they purposely lack much of the variability and unpredictability of inputs in operational situations, and accordingly they are more suitable for skill enhancement than for procedurization. A different

configuration would be to train two interacting individuals from different subsystems—for example, one intercept director together with one interceptor pilot.

Still another configuration is the inclusion of backup modes in an exercise, including transfer of command from prime to backup, especially if there was good reason to believe the prime system would fail. One should train teams in emergency operations—for example, train the team in a nuclear plant control room to cope with a malfunction (as at Three Mile Island) or any team to revert to manual procedures when a computer stops operating. Hostile attack in air defense would be viewed, of course, as an emergency, so almost all SAGE exercising fitted into this category. There was one particular problem. Wartime loads might be so great that their simulation would put so large a demand on the computer and on operational equipment that operational readiness could be jeopardized. There may be a tendency to build into a system for team training purposes a smaller simulation capacity than its operational capacity.

COSTS

One should not underestimate the required effort and cost involved in a simulation-based training program, nor should one overestimate them. Hidden as well as obvious cost factors should be considered for various exercise frequencies for various configurations. As mentioned in the ECCM story, it is difficult to determine optimal exercise frequency. This depends on the exercise configuration, particular stage during the life of the system or occurrence of system changes, skill levels and procedurization needs, and desirability of overlearning. Cost factors include the production of simulation inputs which change from exercise to exercise (for the same team)—hardly a trivial item. Techniques were developed at SDC for building "versatility" into exercise tapes so they could be repeated with the same team but with changes arranged on-site.

TRAINING ENHANCEMENT

I wrote in 1964 that the area of training enhancement was

...rife with careless concepts, confusion, controversy, and uncertainty. Notions of reinforcement have been brought in from animal learning theory, of knowledge of results from human motor skills learning studies, and of non-threatening problem-solving discussions from clinical psychology and from social psychology In spite of the confusion, product enhancement techniques may contribute most importantly to the net training gain, although there are many uncertainties. If as much research and development attention were paid to these techniques as to the technology of simulation, the designer of training subsystems might acquire better guidance.

Skill improvement depends on knowledge of results (KR or KOR), self-administered or given by other individuals or devices. KR can be applied during an exercise or afterwards, to either component or interactional skills. The latter may depend much more than the former on sources external to an operator who may be unable to determine directly the success or failure of his actions involving the actions of others.

Though it is often said that KR should be immediate, this can mean after an exercise (in a debriefing); however, much KR could and probably should be given during the exercise. Even given well after an exercise (with human beings), KR can be effective (due to verbal mediation). The timing should depend on type of task, individual capacity for recall, and especially the nature of the intervening events or activities before the next opportunity to execute similar actions. In any case, it may be more helpful to provide KR just before the next exercise (in a prebriefing) rather than right after the exercise yielding the results, or perhaps at both times.

Exercises can produce procedurization in a number of ways, though exercises, as already indicated, should not be its sole source. Non-exercising experts can procedurize on the basis of their own observations of the exercise or the observations of others. Another method is the critical incident technique of questionnaire and analysis. In the STP, the exercise participants were supposed to discuss afterwards the events and results of the exercise and procedures used and to evolve alternative procedures which might bring better consequences. Considerable knowledge, imagination, and inventiveness would be available if the team contained numerous knowledgeable and inventive members, and the group discussion might serve also as both an instructional medium for those with less knowledge and a source of "team spirit" or "system identification."

As in the case for KR for skill improvement, the often-expressed notion that team discussion for procedurization should follow immediately after an exercise lacks empirical support. Presuming that memory and interest can be revived, a prebriefing prior to the next exercise could be a more suitable occasion for group problem-solving—the results of which could be put into effect almost immediately. New or altered procedures to try out should be discussed in a prebriefing in any case, unless some other method of testing them is used.

Group discussion in a military setting can generate its own difficulties. Potential, inferred, or actual criticism may lead to avoidance or escape behavior, with individuals turning to rationalizations of their actions, blame-shifting, concealment, verbal aggression, or other evasive maneuvers. Not every officer in charge will have or can acquire the skill to prevent or minimize such behaviors, keep the discussion on the track, bring out ideas for emphasis, or forestall verbal free-for-alls. In operational environments, time for discussion must be found and set aside, suitable space made available, and personnel persuaded that after a duty tour discussion is preferable to drinking coffee or going home. The record of the exercise must be presented without finger-pointing and with sophistication, accuracy, and pertinence lest the discussants assail the record-keeping or the equipment and computer programs. Support for procedurization discussions is needed from higher authority so pressures of other duties will not create conflicts and officers will not feel that discussions including enlisted personnel infringe on their authority or responsibility. Within officer-composed teams, the problem of rank may be even more severe—for example, in command staffs.

Different kinds of information are needed from an exercise for skills improvement KR and for procedurization discussions, though the distinction has not always been made clear. Whichever the objective, human observation and record-keeping can have significant requirements. Observers/recorders (preferably military) must be well trained and selected for experience and seniority. They must be equipped with various kinds of aids both for recording data and for informing themselves as

to what has been planned into the exercise—such things as lists, scripts, descriptions, scenarios, and maps. These aids should be human-engineered for effective use. In computer-based systems, observers/recorders may need to have consoles similar to those of the exercise participants, since there may be no central multiviewer display. Simulation inputs may have preprocessed designators (visible only to the observers/recorders) matching planned inputs in scripts or maps, and they should be subject to some control by the observers/recorders (such as removing an aircraft track after a successful interception).

In a computer-augmented system, exercise events including a vast array of performance data can be recorded, reduced, and printed out through programs designed for these purposes. Problems of criteria, measurement, and categorization can develop if the programs were created for testing rather than training. Programs should be versatile enough to serve both KR and procedurization purposes. One type of program can simply record all the processing of input so it can be played back later. Another type records whether events that have been forced by an input program actually occurred and what actions were taken in response; automatic umpiring can determine the success or failure of the actions, though it may be difficult to establish criteria.

Additional techniques to enhance training include carefully controlled competition between teams. Possible rewards are commendations, additional leave time, or even vacation trips. However, it may not be easy to agree on criteria, achieve equivalent inputs for different teams, and prevent attempts to beat the game. One approach would be to have each team design the inputs for every other team, within certain established limits of workload; each team would thus play, in some situations, the role of the enemy—as in war-gaming. Another enhancement technique is to establish standards the team (and its components) should meet, though here as well it may be difficult to formulate the standards and avoid attempts to beat the game. This technique, as with academic examinations, mixes evaluation with training; at least, the participants will probably react to it as though they and their team are being evaluated.

Team exercises are indeed suitable vehicles for both training and explicit evaluation, and hence STP at one point was transformed into SETE (System Exercising for Training and Evaluation), though it was found preferable to have only one of these objectives for a given exercise.

INPUT SELECTION

Some of the quandaries in selecting inputs for training exercises were mentioned in the description of ECCM training—for example, tradeoffs among realism (threat), stress, and difficulty level. Selection and implementation of inputs consume the most time, money, and effort in an exercising program that relies on simulation—and accordingly get the most attention. Simulation technology has acquired a considerable amount of complexity. Simulation packages can be designed with flexibility for alteration for reuse, as noted earlier. More items can be included in a package than can be used in any exercise; various exercise input combinations can be selected on-site from the package. Duration of exercises with the same simulation package can be varied, with different start and cutoff times. Time can be compressed (or expanded). All sorts of unexpected or unlikely—but

critical—events can be introduced, including errors by one's own forces or the adversary, or adversary attempts to deceive one's own team.

SIMULATION EQUIPMENT, MATERIALS, AND COMPUTER PROGRAMS

For team training exercises in an air defense or anti-air warfare system, a substantial technology is required (much of which has been developed) for designing and producing simulation equipment, materials, and computer programs. The equipment consists largely of transducers to convert digital data into simulated radar signals. The materials are primarily the digital data, converted from scripts and scenarios or selected from stores of such data. The computer programs are those that accomplish that conversion or selection in some production facility or modify input data at the exercise site. There is no advantage in going into the details of this technology in this paper, since they are very technical and air-defense-oriented. They are pertinent primarily to team training embedded in a large system. Let me refer you to the report from which this part of the paper is drawn (Parsons, 1964, pp. 35-44). There are, however, a few lessons learned from air defense with regard to simulation production that may apply to other, perhaps simpler, contexts and teams.

Any simulation-based team training will need a capability for producing simulation equipment, materials, and computer programs; in the last instance, an exception might be made of very limited training, training in a non-computer environment, or both. Transducing equipment can consist of operational computers, consoles, and computer programs, various kinds of communication links, or equipment to translate data from one medium to another. Materials can consist of magnetic tapes or discs, various kinds of film, scripts, maps, or other paper, voice recordings, and other media involving new technology.

Computer programs for producing these must take into account (1) operational computers, their programs, and their displays; (2) changes in the above; (3) operational equipment, environments, procedures, doctrine, organization, and communications; and (4) all features of the training effort related to the products. Personnel responsible for operating a production facility must be thoroughly familiar with the system in which the training is occurring and with the team training enterprise. A central facility can have complicated scheduling requirements if its products are used at many locations. It should be viewed as being itself a computer-based system that has to be planned, designed, developed, produced, and operated; without good management, costs will increase, schedules will slip, and initial design will be inadequate. Operational data and simulation technology for checking out the programming of operational computers should be exploited for the production of simulation materials for training exercises. Skilled personnel will be needed. Costs of simulation packages must be estimated as reliably as possible.

To construct computer programs for producing simulation inputs, it may be necessary to assemble a large data base. Personnel will have to gather the data and update them. Mathematical models will be needed for logical processes. Construction of simulation models may be iterative, permitting rapid and extensive modifications, or may try to handle all requirements at the outset (probably in vain). Programmers will need forecasts of kinds and amounts of simulation data. Especial-

ly if the production process becomes extensive, industrial engineering will be needed for materials handling (plant layout, packaging, and storage), production control, inspection, and quality control with feedback from users.

Equipment for team training as for other types of training, including simulation transducers, is likely to include long-lead-time items. Early analyses of needs and compatibilities are essential (though often missing), and as mentioned earlier, equipment must be human-engineered. The operational system designers themselves must understand what is needed for team exercise training in hardware and computer programs, to obviate costly retrofits, gaps, and incompatibilities. Unless there is optimal planning—if team training for air defense is any guide—the simulation materials produced are likely to have significant omissions: (1) some important dimension of the environment (in air defense, for example, aircraft location in altitude); (2) probable mishaps (in air defense, communication outages); (3) contact with effector elements (in air defense, control of Bomarc unmanned missiles and voice communication with data-link-equipped interceptors); (4) contact with coordinated teams (in air defense, data-link communication with Army missiles and command posts); (5) simulated capacity matching possible operational loads (in air defense, hostile aircraft and interceptor aircraft); and (6) "noise"—self-made, nature-made, and adversary-made (in air defense, electronic countermeasures and ground clutter).

Time does not permit me to discuss the other considerations for team training included in my 1964 analysis (some aspects of which have been already noted):

- Supplementary training through instruction and academics.
- Development and installation of simulation exercising.
- Operation and maintenance of simulation exercising.
- Testing and evaluation of simulation exercising.
- Simulation exercising for measurement and evaluation.

CONCLUSION

The ONR-sponsored project being discussed in this workshop is located in the Department of Defense "6.1" category of Research and Development—basic research. One may well ask what bearing the ECCM team training and the problems I have described have on such research. By no means do I suggest that the Rand project should try to develop some team training program for a particular system; that would better fit the 6.3 category of advanced development, or 6.4, full-scale engineering development. Rather, I have tried to suggest some real-world problems that 6.1 research should aim toward helping to solve (as well as some that are mostly managerial or developmental) and to depict an actual team undertaking that may provide one frame of reference (among others) for considering topics to investigate in 6.1 research.

It seems to me there are two ways to generate needs for 6.1 inquiry in team performance. One is to examine advances in related science and technology to see how these may be exploited to make teams perform better. That appears to be the major thrust of this workshop, and I am certainly not against it. In fact, I wish there were more—for example, some attention to behavior analysis and its emphasis on motivational variables. The other way is to look at teams in operation, or actual

team training programs, to find matters that need probing as general problems. A number of these may have suggested themselves in this paper, for example:

- What kinds of interactions and information transformations characterize some teams—from dyads to large teams—and need training?
- What kinds of information feedback can enhance team exercise training, for skill improvement and procedurization?
- How can teams solve problems and develop better procedures (rules) through exercising, and what might such procedures be?
- For simulation-based team training, what levels of information verisimilitude are required, for what objectives?
- What does "noise" information do to team performance, especially when delivered by an adversary, and how can a team deal with it?
- How should difficulty level be established in designing simulated information inputs for team training?

Lest I be accused of behavioral bias, let me point out in closing that all of these questions involve what I understand to be cognitive processes.

DISCUSSION OF SESSION 1

The issue of feedback stimulated a lengthy discussion. Gaines noted that learning involves experiencing failure and using knowledge of how you failed to correct performance. He maintained that the Navy frequently does not permit failure to occur during training. Rather, instructors may correct student operators before they make errors to spare them the adverse emotional effects of failure. Hunt pointed out that the sequence of events leading up to an error can provide useful information. Thus, the instructor might intervene just prior to failure, reducing the threat to the student, while still providing a useful evaluation. Rizzo added that performance in most simulators is rarely measured discretely. Rather, there is a continuum from poor performance through adequate performance to excellent performance. Most operational environments probably share this characteristic.

Thorndyke questioned whether simulator training permits measurement and recording of individual performance while the simulation is in progress. Parsons replied that in the SAGE training program, team members kept logs of their actions and consequences. In addition, individuals received updated information on mission success as the simulation proceeded.

Weiner commented that overall outcome measures were a useful component of feedback, providing a simple standard against which trainees could be evaluated. However, Parsons reiterated the need for detailed process information as well, asserting that outcome measures are not sufficiently specific to guide improvement in particular skills or processes. More detailed feedback enables the learner to direct his attention selectively to areas in need of remediation. Kennedy noted that some simulations lack even rudimentary forms of total-mission-related feedback. He recalled a particular defense simulation in which participants had been unaware that major cities had been bombed.

Edwards expressed concern that simulators are too complicated for many tasks. He suggested that simulator realism should depend on the content of instruction. For example, simplified, schematic simulation environments might suffice in training abstract, intellectually based tasks. Training rigid procedures may require realistic environments, but training intellectual skills in realistic environments may actually depress performance. Parsons agreed, noting an overemphasis on the simulator per se at the expense of the training methodology employed in the simulator environment. Detailed realism is important only for the aspects of the environment that directly influence the skills to be trained. Edwards noted that training for the space program represented the extreme in realism, utilizing the actual operational equipment. However, this situation was unusual since virtually every possible mishap was associated with a response procedure and thus required little real decisionmaking. Weiner summarized the realism issue by distinguishing effectiveness from efficiency. The latter measures the operating characteristics of the team, while the former measures the extent to which the team achieves its objectives. Weiner maintained that realism is important only when it is required to assess effectiveness (e.g., "when the environment bites back"). The discussion concluded with an appeal for a taxonomy of training objectives to govern design decisions such as the realism issue.

III. ORGANIZATION THEORY

HUMAN INFORMATION PROCESSING APPROACHES TO ORGANIZATION AND THE STUDY OF TEAMS: POSSIBLE RESEARCH DIRECTIONS

John P. Crecine

INFORMATION PROCESSING AND MILITARY TEAMS

In assessing the capabilities of teams in military settings (e.g., aircraft of tank crews, command and control system operators, patrol groups) using theories of information processing and decisionmaking in humans and organizations, two basic problems emerge. These problems stem directly from the characteristics of humans and organizations operating in complex environments and carry over to theories designed to deal with these phenomena. And consequently, these problems compound the difficulty of applying theory directly to combat team assessment. The underlying, theoretical problems are:

- Human problem solvers and organizations adapt to task environments. In operation, theories of human information processing and organization are scenario-dependent. Direct application of theory requires a rigorous specification of the task/scenario environment(s). Generalizations are restricted to situations where task environments can be assumed to be similar.
- The mode of adaptation of information processing/cognitive capabilities to the task environment (scenario) is artifactual and not unique. Given the same basic cognitive capabilities and task environments, there is generally a wide range of problem-solving strategies (at the individual level), organizational structures, and processes (at the organizational level) that are roughly appropriate. There is no unique solution to tactical and organizational problems.

This is not to say, however, that such a theoretical approach is not useful and will not produce important insights.¹

Theories, Assessments, and Unique Predictions

Cognitive psychology and information processing theories involving humans (Newell and Simon, 1972) and organizations with human components deal with information processing in particular task environments. Some regularities have emerged in the characteristics of human information processing and problem-solving over a wide range of settings as the result of two decades of experimental and applied research (Slovic, Fischhoff, and Lichtenstein, 1977). In spite of the

¹ As Fisher and Larkey (1979) suggest, information processing theories of human behavior are more likely to be helpful in directing attention to key features of C³I systems and for solving operational problems (or perhaps exploiting such problems in the case of Soviet C³I systems) than in directly providing the means of "scorekeeping" or for direct comparison/assessment of the relative capabilities of two competing systems.

strength of many of the findings, precise predictions of individual and organizational behavior require specification of both the task environment and the way the individual or the organization organize underlying cognitive information processes to perform the task. At the individual level, the mode of adaptation can be referred to as the problem-solving strategy, and at the organizational level, it is the structure of the organization itself and the routines and processes the organization has instituted that represent the mode of adaptation to the task. At the organizational level, the need for coordination leads to a degree of routinization² which in turn suggests that organizational routines are more stable than those employed by individuals in a non-organizational context.

Research Question: If basic research on teams is to be broadly applicable to functional groupings of military personnel in a wide variety of settings, empirically valid ways for characterizing task environments must be developed. How does one characterize all of the functions performed by, say, a tank crew in a combat setting? What are the common dimensions of the tasks performed by, for example, a P-3 squadron in detecting Soviet submarines, an air squadron providing close air support, a tank crew searching out and responding to hostile enemy fire, or a command-control team trying to assess the situation in a particular area of the front? What are the dimensions of task environments that generalize from task to task?

The organization of groups of individuals into large organizations relaxes some of the limitations on individuals as information processors and creates the potential for a wider variety of behaviors than an individual could exhibit. Paradoxically, the requirements for coordination in producing organized behavior are such that one can expect a greater stability in procedures, more routinization of information sources, and more stylized or widely shared "definitions of the situation" than characterize individual-level phenomena. Shared values, common doctrine and practices, and "shared scenarios" also serve as ways of coordinating individual-level behavior informally or indirectly. The very requirements of *organization* reduce the variability of behavior and make it possible to study particular organizations in particular contexts over a period of time.

The next section will briefly outline some key concepts in the theory of human information processing and cognitive psychology. The third section extends individual-level concepts to the organizational level, and the final section attempts to apply these concepts to teams in military organizations and suggests profitable research directions.

KEY CONCEPTS IN COGNITIVE PSYCHOLOGY AND HUMAN INFORMATION PROCESSING

In marked contrast to the assumptions implicit in most social and management science theories of human choice, cognitive psychologists have been centrally concerned with the realism of assumptions concerning human cognition. In particular, there are real limits on the cognitive processes and information processing characteristics of humans. Human cognition is purposive but fallible³; choice behavior

² Processes and individuals linked together in a way so that the behavior of individual parts is somewhat predictable.

³ And certainly not "rational" in the narrow, technical sense that economists, some decision analysts, and management scientists generally assume.

is adaptive but not optimal; search for information and solutions is highly selective, not exhaustive; search is local, not global or thorough. Most of these characteristics reflect humans' adaptations of their own limits on time, memory, understanding, and cognitive ability to the demands of complex problem and information environments. Characteristics of human cognition apply to the ways humans develop an understanding of a complex situation as well as to problem-solving and choice behavior. Some key research findings are as follows:

- People are "unreasonably" (relative to rational models) influenced by concrete, personalized information.
- People judge probabilities by envisioning concrete instances.
- Multiattribute judgments are based on only part of the available information; in making judgments on an issue involving many dimensions, individuals focus on only a very few of the possible dimensions.
- Choice and decision are usually made with respect to an internal "representation," "model," or "definition of the situation." Initially, information is used to construct an internal representation of the situation or, more often, to evoke a previously constructed "definition of the situation." Once constructed or evoked, this internal "model" or representation of the situation is used to interpret other information, generate alternatives, and make choices.
- Experts use the same simplifying strategies as laypeople. Experts often consider even fewer of the dimensions of a situation or problem, relying on internal standards against which ordinal comparisons are made. Experts use fewer dimensions in evoking an (pre-existing) internal representation of the problem; this "definition of the situation" is generally more elaborate than for laypeople, and choice *within* the context of the internal representation is often more elaborate, while being at the same time more tightly constrained by the representation.
- Large amounts of relevant information often lead to inferior and less accurate understandings of situations and to less appropriate decisions. The ability of individuals to simultaneously comprehend large numbers of factors is limited.
- Rather than optimizing (choosing the single *best* alternative), people *satisfice* (choose the first *acceptable* alternative); one doesn't find the sharpest needle in the haystack, only the first one "sharp enough."
- Decisionmakers remember outcomes of past decision situations and, over the course of repeated choices in similar situations, only the habitual decision is retained (affect referral).
- Choices are made on the basis of object attributes (a few relevant dimensions of the problem).
- Memories of (a few) salient features, dimensions, and attributes of a situation are easier to access than are memories based on interval data, tradeoffs involving more than one dimension, or incommensurables; ordinal data are much more easily comprehended by humans than are cardinal or interval-scaled data. There is evidence of a strong bias in decisionmaking toward choices based primarily on those dimensions of relevant alternatives that are commensurable, regardless of their importance.
- Attributes of unchosen alternatives are forgotten.

- For decisions/choices previously made, "higher-level" information gathered in the (past) choice process is remembered; a few salient attributes of past choices are remembered, along with the internal "model" of the situation upon which the choice was made.
- "Lower-level" information is not remembered (from previous choices) but rather is regenerated through "elaborative inferences." Often what is recalled, especially by professionals, is the answer to the question, What should the supporting information have been, given the choice that was made?
- Information acquisition and search are integral parts of the choice process (information environment is important).
- Decisionmaking is often characterized by sequential attention to those dimensions of a problem that are both relevant *and* commensurable between alternatives, using ordinal comparisons on one or two dimensions to eliminate alternatives, working down the list of dimensions until only one alternative remains.
- Humans generally "solve" problems by breaking them up into a series of subproblems, then attacking the subproblems one at a time.

In thinking about the applicability of these concepts to understanding a military organization operated by professionals, in situations important to the individuals involved, and under crisis/stress situations, a number of observations seem relevant. First, professionals seem even more dependent on heuristic procedures for interpreting information and making choices than do inexperienced people. While the professionals' heuristics or "rules of thumb" may be more powerful or functional than those employed by the layperson and may be somewhat more sophisticated, the evidence suggests that heuristics learned through formal training and subsequently rehearsed are both more difficult to change and have the characteristic of attending to just a few aspects of the problem or environment. Professionals cue on a few bits of information and make choices based on relatively few dimensions of the problem/situation. Sophistication in the real world may refer more to the power of the heuristics employed and to the efficiency of search than to the complexity of the situation actually dealt with. Second, decisionmaking, choice, and cognitive activity in a stressful situation exacerbate the tendency of individuals (including professionals) to rely on a few concrete scenarios and on a relatively few pieces of information to evoke the appropriate scenario, which defines the situation and suggests the appropriate course of action.

In most choice situations, whether the decisionmakers are individual laypeople in a complex problem environment or professionals in a complex organizational environment under stress, the decision or choice itself is generally anticlimatic. Most people, most of the time, make decisions by first figuring out "what the situation is" or what scenario they are operating in. The situation or scenario evoked usually carries with it a definition of "appropriate" behavior which guides the decision or choice.

Further, research in cognitive and social psychology and on decisionmaking in complex organizations suggests that managers and professionals systematically select out a few pieces of information in an information-rich environment, base their "definition of the situation" on those few variables, use the resulting model (scenario) or analogy to regenerate or reinterpret related information, and then act

as if the abstraction or model were true. Once formulated, a professional's model of a specific situation is highly resistant to change, even in the face of overwhelming evidence that a particular definition is inappropriate. The role of information in real settings seems to be to help an actor select, from a relatively small number of "models" (highly developed scenarios) or "definitions of the situation," the one the professional will use to guide action and interpret further information.

Research Question: For military teams and crews in various settings and task environments, how do individuals develop/evoke the internal model or "definition of the situation" that they then act upon? Is the "definition of the situation" merely selected from the set of tasks/scenarios developed in training? What informational cues are used to evoke the "definition of the situation"?

INFORMATION PROCESSING BY ORGANIZATIONS: KEY CONCEPTS

One can view organizational decisionmaking and information processing as a linked, coordinated set of individual decision processes. Also linked with individual information processors are machine processors. In some situations, computerized algorithms and other non-human information processing devices are part of the organization.

One of the reasons for creating an organization is to relax some of the constraints on individuals as decisionmakers, information processors, etc. For instance, humans generally approximate serial information processors—one thing at a time. These constraints on time, attention, and expertise are relaxed somewhat when groups of individuals are organized. Although there is a real limit on what organizations can attend to and act upon in an organized way, it is clear that one organization can attend to more than one human can—parallel processing and other kinds of intellectual division of labor and expertise is one reason for organizing. The limits on short-term memory are similarly relaxed; the number of objects or problems that significant parts of the organization can immediately apprehend is much greater than the span for individuals.

While some information processing constraints on individuals are relaxed, the need for coordination introduces others. Both formal (plans and procedures) and informal (shared values, doctrines, scenarios) modes of coordination introduce additional constraints on attention, information flows, and activity. The very need to coordinate individuals and to have communications that mean roughly the same things to sender and receiver tends to force the organization to create stylized representations of the external conditions and realities and, generally, to reduce the variability and "richness" of the information environment in which the organization operates. Partly by restricting the range of communications circulating in an organization, the behavior of the subunits becomes more predictable and manageable and those (restricted) communications circulating internally are more likely to be understood—even if the link with "reality" is tenuous.

Over time, organizations develop structured (and predictable, interpretable) information environments, routines, and standard operating procedures for processing the expected range of information, shared images/values/expectations, and systems of mutual expectations. For the organization, these structures, shared

images, and procedures parallel the stable "definition of the situation" or internal problem representation for the individual. Once created, such images and processes are also highly resistant to change. The routines, sensors, and decisions tend to homogenize a highly variable external environment, to convert unfamiliar information and signals to the familiar, and to convert situational definitions into those sorts of things that look like what the organization knows how to do. An extreme example is that of U.S. military organizations preparing to fight the *last* war, converting particular military threats into situations that the organization has dealt with before.

Research Question: For military teams and crews in particular task environments, how far does the actual task environment have to deviate from rehearsed scenarios before the team's shared "definition of the situation" is modified? How resistant to change are pre-existing definitions of the situation?

Another characteristic of large organizations is partly due to functional divisions of labor and to subsequent socialization of participants around subunit activities and well-defined technologies. Conflicting objectives, technologies, and the like tend to develop among the various organizational subunits. As a result, the overall organization tends to resemble a coalition of partially conflicting interests and subunits rather than a coherent body with perfectly shared values.

Resulting organizational actions are likely to be less than internally consistent or coherent. This is because conflict between coalitional subunits, inevitably, is imperfectly resolved and because there is at least a partial failure to control organizational subunits in a consistent fashion. The larger and more complex the organization, the more divergent the subunit goals and values. The greater the degree of functional specialization, the more the organization must rely on formal procedures to coordinate its activities. In general, organizations do not function in an organized way unless they have rehearsed the organization. Activities involving more than one operational subunit require rehearsal and development of routines.

Research Question: How do teams or crews made of members of different subunits differ from teams drawn from the same subunit in performing coordinated tasks? E.g., command and control operations between different units—air, ground support—or coordination of ASW operations as opposed to tank crews; or maintenance and logistics support operation in a P-3 squadron as opposed to operation of a flight crew.

As part of the rehearsal requirements for carrying on organized activities, there must be formal communication and control mechanisms and a stylized/predictable mode of communications.

Internal cues and information generally dominate in a large organization. Very simple links exist between information in the external environment and the information environment found internal to the organization. The information to coordinate generally dominates the information important for adapting to the external environment.

To understand information flows in an organization, one needs detailed information concerning goals and coalitions of subunits, the routines and procedures/divisions of labor, methods for adapting to changes in the outside world, the nature of the internal (organizationally created) information environment, and the formal modes of coordination and communication. *This is to be contrasted with the ap-*

proach of economists and management scientists: for them, all one needs to know is the task environment. The organization is presumed to behave optimally, given the task.

Research Question: For a particular task/team combination, what are the key informational flows and how do they relate to functional divisions of labor within the team and to the task environment?

SOME THOUGHTS ABOUT EFFECTIVE TEAM-EFFECTIVENESS RESEARCH

Karlene H. Roberts

In considering progress in scientific research and how one selects problems for study, I would like to set my remarks in perspective by citing two statements I have found useful. The first is a remark by Karl Sagan:

Professional scientists generally have to make a choice in their research goals. There are some objectives that would be very important if achieved, but that promise so small a likelihood of success that no one is willing to pursue them There are other scientific objectives that are perfectly tractable but of entirely trivial significance. Most scientists choose a middle course. (1979, p. 58)

While I am less sanguine about scientists not selecting trivial problems than is Sagan, I share his feeling that we do not often choose the most important problems to study—and probably for fairly good reasons in terms of our own personal payoffs.

The second remark was made some years ago in a conversation about the dangers of identifying specific research areas that should be afforded government support. Gene Gloye of ONR said at that time that by highlighting specific content areas for attention, one runs the risk of supporting "research by the yard." No doubt, our experience in the United States with cancer research is an example of supporting research by the yard. We can ask ourselves whether the payoffs for this strategy are anywhere near what we might have hoped them to be by this time. I intend today, however, to tread where angels probably fear to and attempt to identify a list of issues emanating from organizational research that should be attended if we are to make any progress in improving team training, performance, or effectiveness in the future. In so doing, one is bound to uncover some rituals in the field that have proved useful and some that have been deleterious to furthering knowledge.

STRATEGIES AND APPROACHES

A major purpose of this conference is to identify general research approaches inherent in each of the represented areas of research. Identification of approaches and strategies can highlight those that may offer maximum leverage for studying teams and abandoning those that appear to lead us increasingly to spin our wheels. Thus, I will begin with a few remarks about major approaches and strategies I see as functional for thinking about team research. I will then mention several things we do not do that we should be doing. Finally, I would like to draw attention to several approaches that ought to be discarded entirely because they are misleading.

The most obvious characteristic of organizational research is the eclecticism of the participants in the endeavor. We represent a variety of basic disciplines in the

social sciences (from human engineering to industrial and social psychology, sociology, economics, anthropology, and even political science), and our concepts, paradigms, and language are borrowed freely from among disciplines (Roberts, Hulin & Rousseau, 1978). While this has led to an embarrassment of riches in approaches, concepts, and meanings attached to those concepts, and a consequent lack of understanding about just what we should be studying, it has one distinct advantage. Organizational researchers are possibly more used to dealing with issues that transcend disciplines than are other behavioral researchers, with the result that while some of us find it necessary to hang on to our defensive posturing, others are working to amalgamate ideas emanating from several disciplines. In fact, one contribution of organizational sociologists was to bring to our attention the importance of thinking about construct homology (similarity of structure of constructs due to common ancestral form) across organizations, groups, and individuals (Hannan, 1971; Robinson, 1950, etc.).

Second, the majority of organizational researchers are committed to studying responses "in situ." This is an asset for two reasons: questions of the external validity of our research are reduced, and the complexity of phenomena thought to influence responses can best be captured in real organizations.

Third, theoretical attention is being given increasingly to the environments in which responses occur. This clearly represents a pendulum swing from our recent past in which the majority of emphasis was on individual differences. However, one should not be fooled by this. While our rhetoric speaks to the potential importance of environmental impacts on individuals or teams or organizations, our research is not, by and large, inclusive of environments. There is a danger that it will become inclusive of environments, while the basic units of interest, whether individuals, teams or organizations, in those environments will become forgotten. Our analytic techniques are sufficiently developed that we need not forget either.

Similarly, organizational theorists call for the inclusion of time in research designs. However, there are few suggestions about how to do this (as counterexamples, see Hulin and Rousseau, 1980; Roberts, Hulin, and Rousseau, 1978), and embarrassingly fewer instances in which such an approach have been attempted.

In many ways it is easier to point the finger at what we are not doing that we should be doing than to recognize those strategies that appear to be leading to some progress. With great attempt at self-restraint, however, I would like to mention three glaring problems I see in organizational research that should be avoided in future research on teams.

In much of the research I referee for potential funding or for journal publication, the data analysis and techniques are overly sophisticated given the quality of the data or the maturity of the underlying question(s). In fact, I have a colleague who is fond of saying that most of the questions we ask deserve no more sophisticated analytic treatments than that afforded by a chi square. I do not think this suggests that we should revert to primary data presentations in the form of frequency counts. I do think it suggests the need to work harder on problem conceptualization—possibly to fund that work separate of research implementation with the promise that the ensuing research will be funded if conceptualization meets some specified standard. Consistent with this, our measurement spectra must be broadened. The almost total reliance on questionnaire responses must come to an end.

A second strategy we must pursue with more effort is that of ensuring consistency in levels of analyses in underlying models, the data collected to test them, and

the analytic strategies used to make those tests (see, for example, Roberts and Burnstein, 1980). We have been reminded a number of times of the need for consistency across these three aspects of the research process since Robinson's (1950) seminal article on the ecological fallacy and the presentation of the Coleman report (1966). It would be embarrassing to many of us if I were to detail the large bodies of organizational research in which there are glaring errors of this nature—errors that, by and large, have never been recognized, much less corrected. Often the data collected cannot possibly address the question posed, and one wonders whether the researcher wanted to answer the question posed, or the question suggested by the data—if the data suggest any questions at all.

A third activity we need to engage in, particularly in team research, has to do with considering the homology or lack thereof among constructs developed by researchers who take different perspectives of organizations. For example, researchers are already assiduously engaged in thinking about team effectiveness. Is team effectiveness the same as aggregated individual performance, or does it represent some different construct? Is team morale the equivalent of aggregated individual satisfaction? The question of homology is extraordinarily important in an area in which little is known and in which researchers are bound to borrow constructs from other areas, probably areas developed by students who took different organizational perspectives.

I would like to quickly mention two approaches that should be discarded altogether, for reasons that are probably obvious. We need no more bivariate investigations and no more studies consisting solely of correlated response error. Within-person or cognitive models—the only kinds of models that lend themselves to this kind of data collection—can usually be tested using some multimethod approach.

In summary, then, I have identified three characteristics of organizational research that determine our strategies and might benefit team research. Our eclecticism and general commitment to in situ research should result in rich research designs. Our lip service to attending impacts of environments and time on responses, if operationalized, should extend our research domain and give a process flavor to it. I have suggested that we work harder on conceptual development and measurement so they might match in complexity the available analytic technology, and that we worry about consistency in concepts, measurement, and analysis. Returning to conceptualization again, I have indicated the need to be concerned with construct homology or the lack thereof, particularly when constructs are freely borrowed across disciplines. Finally, I have suggested that we entirely eliminate bivariate and correlated response error investigations.

DIRECTIONS FOR FUTURE RESEARCH

A second goal of this workshop is to consider "directions or paradigms" for future research. One might also think about whether the payoffs for pursuing particular issues are important or trivial and whether the time to payoff might be long or short. One criterion for assessing payoff triviality is expected value in terms of applications. Another is reduction in some kind of cost. Other criteria also exist. All in all, however, estimating the importance or triviality of a potential finding, particularly in the long term, is a guessing game. Estimating the relative time to payoff is a little easier and possibly a little more accurately done. Here one can

consider how much attention has already been devoted to the problem and the degree to which measurement strategies exist for addressing the problem area.

I would like to focus on four research areas I think will produce findings that will not be trivial or trivial squared. First, important in choosing any set of variables to study is the choice of dependent variable(s) of interest. I note that Rand and perhaps the Defense Science Board Task Force on Training Technology used the words performance and effectiveness somewhat interchangeably. They are not interchangeable; this is all too well known by ONR, who supported a set of important discussions on organizational effectiveness (Goodman and Pennings, 1977). In that set of papers, John Campbell (1977) makes the point that one needs to consider a systems and an outcome approach to organizational effectiveness simultaneously. This statement is no less true for teams, whether they be temporary or of long duration.

Since numerous performance measures already exist, the weight of the effort should be placed on other aspects of effectiveness, particularly those relevant to teams as systems. Attention, too, should be given to the development of measures that allow for comparisons across teams performing different functions. Hackman and Oldham (in press) have made some progress in developing a model of final and intermediate criteria of work-group effectiveness. It might be profitable to ask if there is anything infectious about effectiveness. Does it spread? Actually, one cannot move very rapidly to the independent variable side of the team problem until questions of effectiveness have been better addressed. Since the industrial psychologists among us, in particular, have long considered such criteria problems (although generally in terms of individuals), it seems likely that organizational researchers could develop systemic and outcome measures of team effectiveness that could be used in concert. Because attention has been given to similar problems for individuals and for organizations and because some concepts and methods exist, such development in team research should take an intermediate rather than long or short time.

Even if we understood team outcomes better than we currently do, it would be difficult, particularly in military organizations, to know how to select or train team members unless we had a better grasp than we do of differential technologies that operate in teams, particularly interlocked teams. Resupply, tactical, communication, firing, and other systems must interact. One question of effectiveness has to do with their technological compatibility.

Organizational researchers commonly define technology as a transformation process (Billings, Klimoski, and Breaugh, 1977). Rousseau (1979) extends on this focus, indicating that organizational technology has three major phases: input, conversion, and output. Of these three phases, conversion has been studied almost to the exclusion of input and output. Input can be subdivided into characteristics or attributes of material, information, or people, and into control, or forces influencing the availability and distribution of inputs (stockpiling, screening, etc.). Output can also be divided into characteristics and control (factors influencing the quality and quantity of output released). Rousseau proposes a framework for assessing organizational technology and relating it to organizational environments and traditional managerial concerns such as links between technology and structure. Team researchers, too, should adopt a more comprehensive view of technology and examine input-transform-output processes within teams and across interlinked teams. Since there are problems in both construct development and measurement (the

area, after all, has been central to organizational research only since Woodward's work (1958, 1965)), this area is going to require some time to develop. Nevertheless, we should push on.

Dictionary definitions of teams (*Compact Edition of the Oxford English Dictionary*, 1971) discuss yoking or harnessing together team members. In a very real sense, how they are yoked together and how they are yoked to other organizational elements are important. These things are important if we are to assess technological compatibility across units, or get a handle on how to design teamwork. Research on communication in organizations has focused on information flow both within work groups and within organizations (Porter and Roberts, 1976; Roberts et al., 1974). How information is used for decisionmaking has not been examined. Research on organizational decisionmaking is limited because it has focused primarily on decision events and decisionmakers largely in isolation from organizational environments. In addition, most of this work has been conducted in laboratories which, at the very least, reduce the abundance of possible influences on communication and decisionmaking that can be examined. The time has come for these two research traditions to come together in field investigations.

O'Reilly and Anderson (1979) suggest that future research in this area focus on the influence of contextual factors (such as structure, incentive systems, power, and the development of norms), both on information acquisition and information processing by decisionmakers. For example, how do commanding officers select, process, and use information in naval landing exercises? How is the joint development of weapons systems by NATO members to be accomplished with maximal effectiveness in a milieu in which a wide variety of structural, power, incentive, and cultural processes operate?

Another aspect of team functioning that should be examined in communication terms has more to do directly with the connectedness issue. Perhaps some teams operate most adequately when people are minimally tied together. Thus, one might ask questions not about how to tightly compose teams, but rather about the functionality of various degrees of decomposition. One might be able to improve conceptualization of and even operationalize such notions as the strength of weak ties, near decomposability of systems (Simon, 1969), or loose coupling (Weick, 1976), in addressing such questions.

For some research investments in the communication decisionmaking area, one can expect rapid payoff. Many questions worth asking can be addressed using sociometric approaches, a set of approaches currently technically well developed but theoretically poverty stricken. Buttressing available technology with some theory will probably lead to some quick answers. In areas that require detailed conceptualization from both communication and decisionmaking, the decisionmaking issues have more to offer; payoffs will be delayed by the necessity to articulate better communication theory and amalgamate it with existing decisionmaking theory.

If one had better information in the areas already mentioned, the appropriate design of teamwork would be easier to think about. One finds it difficult to think about team selection, training, and effectiveness in the absence of work design. While issues concerned with structural determinants of organizational effectiveness are practically dead, process approaches to organizational design may provide

insights to team designers. This literature is a gold mine of ideas, most of which have not been translated into empirical research (Kilman, Pondy, and Slevin, 1976). As a caution to overdesigning teams and systems, one might think about self-designing systems. In a discussion of this issue, Weick (1977) uses as an example the Apollo 3 program in which the Skylab crew activities were so overly designed that the crew ultimately went on strike—our first work stoppage in space.

Research done in recent years on task design (Hackman and Lawler, 1971; Hackman and Suttle, 1977) might offer some ideas about how to proceed. However, I would urge caution, not only because of potential problems in homology of constructs developed to think about tasks individuals do being applied to teams, but because there are substantial conceptual and measurement problems in this work (for example, see Schwab and Cummings, 1976). Hackman (1979) presents a thoughtful argument about how to think about work design in groups that offers some interesting suggestions about group-task structure, work-team composition, and the development of work-group norms. He notes that there is a diminution of work on "juicy interpersonal process issues" and that these issues should be put aside until we examine more carefully underlying causes of interpersonal dysfunctions. Payoff in the area of teamwork design is probably a long-term proposition. So many ideas are available in the macro organizational literature that it will take time to sift through and identify the most fruitful ones and to develop ways to operationalize them. The micro literature is characterized primarily by testing one model, using an extremely limited kind of measurement.

In sum, I have identified four areas in which it appears research effort should be supported in terms of furthering our knowledge about team effectiveness. Before further steps can be taken, one must give considerable attention to the criteria problems. Technology, communication, and decisionmaking probably contribute to team effectiveness and the effectiveness of interlocked teams. Finally, all of this seems in vain if we are to ignore problems of design.

Obviously, the best research programs in any area will be designed such that their findings can later be integrated with those from other areas. To this end we must encourage discussion among researchers who will be examining issues of team effectiveness from a variety of different perspectives.

Let me end with just two cautions:

1. We should not be deflected into thinking that what is needed is the study of group influences on individuals. Many people think social psychologists study groups. Most frequently they study individual responses in groups. We need investigations in which the basic unit of consideration is teams and responses unique to teams, not in which the basic unit of interest is simply aggregate individual responses.

2. We should not be seduced into thinking that because a problem is described in information processing terms it is automatically worth studying. Recently, there has been in organizational research a move to discuss numerous kinds of problems as information processing problems, from motivation (Salancik and Pfeffer, 1978) to commitment (Staw and Fox, 1977) to task design (O'Reilly and Caldwell, 1979). This probably parallels the move in social psychology away from dissonance and toward attribution models (Ross, 1977). While some of this work may be valuable, it is doubtful that a wholesale move to describing everything in information processing terms is going to do much for us other than change the labels we use.

DISCUSSION OF SESSION 2

Nilsson commented that a team that hierarchically subsumes several smaller teams will not necessarily behave according to the same principles as the smaller teams. Crecine agreed that equating teams at different levels of the hierarchy could be misleading.

Hunt shifted the topic to another aspect of team performance: time demands of the task. He noted that teams whose tasks require instantaneous reaction to changes in the environment represent only one type of team. Teams without such temporal constraints may have very different psychological demands. According to current promotion policies, however, team members typically are promoted to time-independent teams by performing well on a time-stressed team. Hunt warned that research must not ignore time-independent teams simply because they are less amenable to laboratory investigation.

Menchik pointed out that time-stressed teams have conflicting requirements for flexibility and rapid response. He suggested that one might learn about coordination by studying how a team controls its internal communications and how it develops a shared language. Job rotation may help develop a shared body of experience and hence may facilitate effective communication. Whether this is the best method for developing coordination skills, however, is a question for research.

King returned to the discussion of team characteristics. He maintained that the Navy needs to know how to distinguish teams and "partition the variance" in their performance. Hunt raised the related requirement of establishing criteria that dictate when to assign a problem to an existing team and when to establish a new team. Weiner asserted that broad commonalities in behavior exist across team types. The Joint Chiefs of Staff (a high-level decisionmaking team) and an infantry battalion (a highly proceduralized team) deal with similar questions: "What is . . . ?" questions (i.e., requests for status reports) and "What if . . . ?" questions (i.e., contingency planning). Weiner suggested that research might treat the task environment sufficiently abstractly to capture the similarities across team types. Parsons claimed that these similarities are artifacts of the verbal nature of both infantry and JCS tasks. Weiner replied that an infantry battalion performs many nonverbal tasks, but nevertheless their communications and functions resemble those of more "verbal" teams at an abstract level.

Crecine shifted the discussion to the topic of organizational policies. He noted the differences between U.S. and Soviet military incentive structures. For example, job rotation, designed to encourage flexibility, is an organizational policy in the United States but not in the Soviet Union.

Focusing on the flexibility theme once again, Hunt noted that an individual's flexibility in considering alternative problem solutions declines markedly with age. In some circumstances, the longer a person remains in the Navy, the poorer his decisions may become.

Collins returned to the issue of group structure, requesting evidence that the determining structural organization interacts with task characteristics in performance. Roberts replied that early organizational research had hypothesized such an interaction, but that the evidence in the literature was equivocal. It was noted

that the most effective structure may not be the structure most conducive to high morale. Apparently, evidence for the proposition that "happy workers are good workers" is not compelling.

IV. SMALL GROUP PROCESSES

PROCESSING ARGUMENTS AND FORMING OPINIONS IN GROUPS

Eugene Burnstein

We have been concerned with the polarization of opinions in stable, face-to-face decisionmaking or problem-solving teams in which status differentiation and specialization are minimal. Our approach, called persuasive-arguments theory, leans heavily on a fairly longstanding tradition of work on informational influence as well as on current research in cognitive psychology. Group members, thus, are viewed as information processors and groups as information processing systems. More specifically, we assume that when a group member evaluates (or reevaluates) alternative positions, points of view, courses of action, etc., say, *J* versus *K*, he generates arguments, namely propositions describing the attributes of concepts *J* and *K*. We assume that there exist in memory relevant knowledge structures, that is, an organized pool of culturally given arguments speaking to each alternative. In order to judge the relative merits of these alternatives, the person assembles arguments from this pool and integrates them into an opinion. Up to now we have focused on several properties of such arguments: their number, availability or accessibility (the probability of their coming to mind), direction (pro-*J* or pro-*K*), and persuasiveness. In outline form the theory says that (1) when the preponderance of arguments in the pool favors a particular alternative, the average prior opinion reflects the direction and magnitude of this preponderance; (2) further thought or discussion on the issue leads to the polarization of opinion toward the alternative that elicits more and/or better arguments; and (3) the extent of polarization will depend on whether the initial sets of arguments that members access overlap with or exhaust the large pool. This implies that polarization will occur to the degree that the average member's initial position is based on a less than complete processing of the information available in the argument pool. The latter can be due to real ignorance (e.g., some people may never have encountered a particular idea) or to a frailty of thought (e.g., some people may forget an idea momentarily). Whatever the reason, under these conditions there remain many important arguments which have not yet come to mind when the person begins to rethink the issue; when several individuals discuss the issue with each other, the arguments which have come to mind are only partially shared.

A persuasive-arguments analysis of *discussion* effects may be illustrated with a simple example. Consider a choice in which the culturally given pool contains six pro-*J* arguments, *a*, *b*, *c*, *d*, *e*, and *f*, and three pro-*K* arguments, *l*, *m*, and *n*. One of several distinct outcomes would be expected, depending on the distribution of arguments among members. Suppose all three of our discussants had thought of the same arguments. In this case, their prior attitude toward *J* would be identical and discussion would produce no change. On the other hand, if *a*, *b*, and *m* were retrieved by one discussant, *c*, *d*, and *m* by the second, and *e*, *f*, and *m* by the third (i.e., if each had assembled different pro-*J* arguments, but the same pro-*K* arguments), then, although they again hold identical prior attitudes, the discussion would produce marked polarization toward *J*. Finally, polarization toward *K* would

be predicted if one member had retrieved *a*, *b*, and *l*, another *a*, *b*, and *m*, and the third *a*, *b*, and *n* (i.e., if each had initially thought of the same pro-*J* but different pro-*K* arguments).

THE PERSUASIVENESS OF AN ARGUMENT

After our first few studies on polarization, it was clear that in describing persuasion we had to take into account somehow the fact that group members generate (infer) additional arguments in thinking about those presented during the discussion. The opinions they eventually form, therefore, reflect both internal and external sources of information. Indeed, an understanding of how individuals construct, that is, analyze and elaborate upon, arguments seemed a key to information influence. Nonetheless, until our more recent work, the concept of persuasiveness received little theoretical analysis along these lines. The neglect, however, was benign. First, in order to predict polarization, there was no need to explain *why* an argument was convincing or to define this property in more than operational terms (e.g., according to judges' ratings of persuasiveness). Second, such an explanation was difficult, given what was known (or, rather, what we knew) about human information processing. However, we did suggest at a less formal level that persuasiveness depended on the implicit meaning of an argument, that is, whether the argument seemed to suggest additional ideas and, of course, whether these ideas were sound (Burnstein et al., 1971; Vinokur, 1971; Vinokur and Burnstein, 1974).

In exploring different ways of looking at the relationship between persuasiveness and semantic content we were struck by two findings. One was that opinion change is only weakly related to retention of the explicit content of an argument (e.g., Greenwald, 1968). The other was that opinion change is strongly influenced by the ideational embellishment that occurs either in comprehending an argument (e.g., Brock, 1967; Greenwald, 1968; Keating and Brock, 1974; McGuire, 1964; Petty and Cacioppo, 1977; Petty, Wells, and Brock, 1976) or in merely thinking about the issue (e.g., Burnstein and Vinokur, 1977; Tesser, 1976; Tesser and Conlee, 1975). On the basis of these data we thought it would be useful to view persuasiveness as a by-product of semantic elaboration processes which establish associations between the proposition contained in an argument and those represented in prior knowledge. As a result of such associations, hitherto inaccessible propositions come to mind and are integrated to form an opinion.

According to this analysis, an argument is persuasive because it activates other arguments that otherwise would not have been made. For example, one of the Choice Dilemma items widely used in research on polarization (Kogan and Wallach, 1964) has the group decide whether someone should attend a university that is highly distinguished but extremely difficult, so that a sizable number of students fail to receive their degree, or a university that is considerably less distinguished but easy enough that virtually every student who enters receives his degree. In group discussion of this problem there are a few arguments that are extraordinarily cogent. One of these is: "He can always transfer." This proposition is persuasive because once it is asserted an important inference is automatically drawn: "Going to the better school and flunking is not much worse than going to the lesser school to begin with." Similarly, another item involves recommending to a prosperous

domestic corporation that it invest either at home or in a foreign country. In the latter case, the rate of profit would be high but the safety of the investment cannot be assured because the political situation is such that there is a probability that foreign holdings might be nationalized with no compensation. Here a rather persuasive argument is: "Think of the tax write-off if they're nationalized!" In both cases the argument ("He can transfer" or "Think of the tax write-off") adduces new information that increases the attractiveness of the uncertain alternative and affects (polarizes) opinions in the group (Vinokur, Trope, and Burnstein 1975). Notice also that each argument is itself an inference based on the members' pre-existing knowledge about "choosing between universities" and "foreign investment situations."

In the remainder of this paper I want to discuss our most recent research on informational influence and argument processing. First, I will present some findings on group depolarization. Then I would like to describe an experiment in which an argument-processing analysis is applied in an Asch-like conformity situation.

DEPOLARIZATION

It is well known that in the typical polarization situation, attitudes can also depolarize (converge) and become less extreme. For example, about a quarter of the standard choice-dilemmas are neutral in respect to polarization, that is, discussion leads initially riskier group members to become more cautious to the same extent that it leads the initially more cautious group members to become riskier. Thus, depolarization alone is observed. Although this pervasive effect seems to be considered theoretically unimportant, it can be demonstrated that in certain circumstances the two types of theories in question, comparison and persuasive-arguments, make grossly different predictions regarding the relative magnitude of polarization and depolarization. Let me now describe these circumstances.

In an analysis of neutral (non-polarizing) choice dilemmas (Vinokur and Burnstein, 1974), we found that these items depolarize (and do not polarize) because their argument pools contain a similar number of partially shared persuasive arguments in favor of the certain (cautious) and uncertain (risky) courses of action (also see Vinokur, Trope, and Burnstein, 1975). The identical state of affairs may also occur on occasion with items that are known to polarize. Consider an item that typically shifts toward *K*. According to persuasive-arguments theory, the average individual will have access to more pro-*K* than pro-*J* arguments, and hence, his prior attitude will be pro-*K*. Suppose, however, there is an unusual group in which half of the members are pro-*J* (and thus, for the moment, have more pro-*J* than pro-*K* arguments) and half are pro-*K* (and thus, for the moment, have more pro-*K* than pro-*J* arguments). Even though discussion typically leads to polarization toward *K*, persuasive-arguments theory must predict that depolarization will occur in this group—the pro-*J* subgroup will become more pro-*K*, the pro-*K* subgroup more pro-*J*, and there will be relatively little polarization. Furthermore, if the proportions of pro-*J* and pro-*K* arguments in the larger pool are known, even roughly, then an estimate can be made of which subgroup will depolarize the most. For example, on dilemmas that usually polarize toward risk, the number of pro-risk arguments in the pool is larger than the number of pro-caution arguments. It

follows that in discussion, the pro-risk members should be able to marshal a greater number of additional arguments supporting their position than the pro-caution members. As a result, the pro-caution faction should shift (toward risk) more than the pro-risk faction (toward caution). A similar line of reasoning would predict that pro-risk members will exhibit a greater shift (toward caution) than pro-caution members (toward risk) on dilemmas which typically polarize in a cautious direction. Finally, recall that on neutral dilemmas, the proportion of pro-risk and pro-caution arguments in the larger pool is similar. According to persuasive-arguments theory, therefore, pro-risk and pro-caution should depolarize to the same extent.

We recently performed two experiments to test these ideas. In the first experiment, groups of six individuals discussed the standard choice-dilemma items. Each group was composed of two attitudinally homogeneous subgroups, with the position taken by members of one subgroup being markedly different (and opposite) from that taken by members of the other. This state of affairs either was not mentioned at all (subgroups not salient) or was explained in detail prior to discussion and kept salient throughout by an appropriate seating arrangement and the assignment of highly visible labels (subgroup salient). The second experiment also involved group discussion. However, the items, compared to those used in the first experiment, were highly variegated.

The results of the first experiment were unambiguous. Since subgroup salience did not significantly affect the mean amount of polarization (shift) or the mean amount of depolarization on any single item or on all items pooled, the data from the two experimental conditions were combined. These data are summarized in Table 1. As can readily be seen, the creation of distinct subgroups with divergent points of view did not affect the shifts typically found with such items. That is to say, there was a significant shift toward risk on risky items, a significant shift toward caution on the cautious items, and no reliable shift whatsoever on the neutral items. It is quite important to keep in mind, however, that polarization in this as well as in the second experiment is comparable to that observed in past studies only at the gross level of group averages. The pattern of changes that underlies the gross effect is rather different. In past research, polarization was due to the more moderate members becoming more extreme. In our experiment, the two subgroups contain members who are equally extreme to begin with and in every case following discussion they converge, i.e., become less extreme. On converging, however, one subgroup depolarizes more than the other (e.g., on items known to shift in a cautious direction, the subgroup that initially is extremely cautious will move in a risky direction and the subgroup that is initially extremely risky will move in a cautious direction, but the latter subgroup will move more than the former). Thus, while both subgroups always depolarize and never polarize, the average position of group members will appear more extreme—which is a long-winded way of saying that, because of the differential rate of depolarization, the average post-discussion preference will be more cautious on cautious-shifting items and more risky on risky-shifting items than the average pre-discussion preference.

The depolarization effect itself is sweeping. The initial mean gap between subgroups, which amounted to 4.90 scale points (on a ten-point scale), decreased by half to a mean of 2.46 scale points. Hence, a short discussion of six minutes halved the large difference between two subgroups whose members were clearly aware of their being in opposition. Furthermore, on each item the magnitude of depolariza-

Table 1

MEAN CHOICE AND MEAN GAP BEFORE (I) AND AFTER (II) GROUP
DISCUSSION, MEAN SHIFT AND MEAN DEPOLARIZATION FOR
CHOICE DILEMMA ITEMS

Item Type	N	Mean Choice	Mean Shift ³	Mean Gap ⁴	Mean Depolari- zation ⁵
<u>Risky Items</u>					
A	14	I. 4.71	-1.21***	I. 5.09	3.09***
		II. 3.50		II. 2.00	
B	16	I. 5.03	-.58***	I. 4.31	1.95***
		II. 4.44		II. 2.35	
C	19	I. 4.58	-.70**	I. 4.79	2.59***
		II. 3.88		II. 2.19	
K	19	I. 5.25	-.78**	I. 4.47	2.08***
		II. 4.47		II. 2.38	
<u>Neutral Items</u>					
D	14	I. 6.59	.18	I. 4.90	1.92***
		II. 6.77		II. 2.97	
H	18	I. 6.22	-.10	I. 4.79	2.01***
		II. 6.35		II. 2.77	
<u>Cautious Items</u>					
E	19	I. 6.22	.98***	I. 5.96	3.36***
		II. 7.21		II. 2.59	

¹I = Mean choice before discussion (initial choices).

²II = Mean choice made after group discussion in private.

³Negative values indicate shifts toward risk. Positive values indicate shifts toward caution.

⁴This mean indicates the absolute difference between the two subgroups (i.e., the gap) before (I) and after (II) group discussion.

⁵Mean depolarization indicates changes in the mean gap between the two subgroups following group discussion. Positive value indicates a *decrease* in the gap, i.e., depolarization. Negative value indicates an *increase* in the gap, i.e., polarization.

**
 $p < .01$.

 $p < .001$.

tion of the part of one subgroup relative to that of the other subgroup was just what would be anticipated on the basis of persuasive-arguments theory.

Persuasive-arguments theory implies that if each individual accesses and integrates *all* the relevant information before choosing (i.e., if the argument sample he or she generates exhausts the larger pool), the impact of discussion will be nil. Nothing anyone may say can change that person's mind because he or she has already taken everything into account. In other words, discussion will produce a shift in attitude only when the relevant information is partially (rather than completely) shared (Kaplan, 1977; Kaplan and Miller, 1977; Vinokur and Burnstein, 1974; Vinokur, Trope, and Burnstein, 1975). In our second experiment we made a reasonable guess as to how information is shared in respect to different kinds of issues. If the guess is correct, then we have distinguished those issues in which discussion will have a pronounced effect on attitudes from those in which it will not. Later we shall present some evidence, gathered as an afterthought, in support of this guess. Now let us simply describe the reasoning behind it. It is not far-fetched to assume that the commerce individuals have with an issue (the extent to which they read, think about, and discuss the matter) is a rough index of *information* and *interest*. If individuals *often* read, think about, and discuss a topic, they will know much of what there is to know about it. Hence, their knowledge is likely to overlap considerably. On the other hand, if individuals *rarely* read, think about, and discuss a topic, they will know only a small part of what is to be known. Hence, their knowledge is likely to overlap very little. Thus, the more popular the topic, the more widely shared the arguments. And the more widely shared the argument, the weaker the impact of group discussion.

Given this line of reasoning, we then conjectured that our population was likely to have the most commerce with certain matters of social value (e.g., Is capital punishment justified?), a moderate amount with certain matters of personal taste (e.g., Is attending a classical music concert more interesting than visiting an art museum?), and least commerce with obscure matters of fact (e.g., How far below sea level is the town of Sodom?). According to persuasive-arguments theory, then, if attitudinally divergent subgroups discuss these topics, depolarization will be greatest for matters of fact, moderate for matters of taste, and least for matters of social value.

Except for the items discussed, our second experiment was identical to the previous one. Again the results indicate that the manipulation of subgroup salience had no significant effects on polarization or depolarization. The data from the two experimental conditions were therefore combined and are shown in Table 2.

As can readily be seen, one value item and two factual items exhibited statistically significant polarization; the two taste items gave no shifts. Consistent with persuasive-arguments theory, the strongest depolarization was exhibited by the factual items, a moderate amount by the taste items, and the least depolarization by the value items. Depolarization on the choice-dilemma items fell between the factual and the taste items. The proportional reduction in the initial (pre-discussion) gap was .81 for the factual items, .49 for the dilemma items, .35 for the taste items, and .18 for the value items.

Finally, remember that our predictions regarding differences in depolarization over items assumed corresponding differences in knowledge and interest in the issues. We made a modest attempt to assess the validity of this assumption by asking another group of individuals to perform four rank orderings of the seven

Table 2

MEAN CHOICE AND MEAN GAP BEFORE (I) AND AFTER (II) DISCUSSION,
MEAN SHIFT AND MEAN DEPOLARIZATION FOR FACTUAL, TASTE,
AND VALUE ITEMS

Item Type	N	Mean Choice	Mean Shift ³	Mean Gap ⁴	Mean Depolari- zation ⁵
<u>Value Items</u>					
A	29	I. ¹ -.06 II. ² .05	.12	I. 5.53 II. 3.86	1.66***
B	20	I. -.28 II. -.70	-.41**	I. 7.23 II. 6.73	.50
<u>Taste Items</u>					
C	22	I. .18 II. .19	.01	I. 6.22 II. 4.27	1.95***
D	15	I. .23 II. .38	.15	I. 5.84 II. 3.55	2.28***
<u>Factual Items</u>					
E	23	I. 7.25 II. 7.87	.62	I. 7.89 II. 1.13	6.76***
F	18	I. 9.18 II. 8.16	-1.02**	I. 6.88 II. 1.20	5.68***
G	23	I. 8.89 II. 11.47	2.58***	I. 7.72 II. 1.95	5.76***

¹ I = Mean choice before discussion (initial choices).

² II = Mean choice made after group discussion in private.

³ Negative values indicate shifts toward risk. Positive values indicate shifts toward caution.

⁴ This mean indicates the absolute difference between the two subgroups (i.e., the gap) before (I) and after (II) group discussion.

⁵ Mean depolarization indicates changes in the mean gap between the two subgroups following group discussion. Positive value indicates a *decrease* in the gap, i.e., depolarization. Negative value indicates an *increase* in the gap, i.e., polarization.

** $p < .01$.

*** $p < .001$.

items that were used in the second experiment. The rankings were made in respect to (a) "How many ideas do you have about the issues? When you try to make up your mind about these issues, can you think of many ideas or only a few ideas?" (b) "How often have you read about these issues?" (c) "How often do you discuss these issues?" and (d) "How interesting are these issues?" The rankings on all four dimensions were perfectly correlated, with strong interjudge agreement, demonstrating that our subjects had the most commerce with and interest in matters of social value, a moderate amount regarding matters of personal taste, and the least regarding matters of fact.

INFORMATIONAL INFLUENCE AND CONFORMITY

Argument processing sometimes may seem irrelevant in the light of common sense. After all, there are well known group situations in which shifts in attitude occur when there seems to be virtually no information available for members to think about (or discuss) in constructing their attitude (e.g., Sherif's autokinetic paradigm), or in which the available knowledge appears to argue for an attitude quite different from the one actually observed (e.g., Asch's line-judgment paradigm). Of course, we are setting up a straw man by assuming that argument processing must be conscious or group members must have access to these operations. Informational influence cannot be dismissed just because introspection says that informational events seem feeble or difficult to access. Such introspections, however, touch on an interesting problem and I would like to speculate about this problem in the context of the Asch conformity situation (Asch, 1951; 1956). Whether or not Asch's line-judgment task is in fact "unthinkable" (and "undiscussable"), we will assume for the moment that argument processing does occur but is difficult to tap directly. This assumption has testable implications.

A number of situation factors guide (bias) group members in processing arguments. For instance, knowing that a disagreement with the majority will be made public may change the pool of relevant arguments and lead the individual to consider implications that would never have entered his mind were the disagreement to remain private. Indeed, in terms of persuasive-arguments analysis, some widely used experimental procedures for changing attitudes may actually do so because they introduce new information which inadvertently redefines the universe of relevant arguments for the subject. As a consequence, implications that were inaccessible or had no bearing on the issue become accessible or pertinent. Redefining the pool, in the sense of enlarging it, may be critical for polarization when there would be few arguments available otherwise (e.g., when the person is uninformed about the issue). Suppose a group is dining and one member is asked to state his opinion about something with which he is quite unfamiliar, say, kumquats. Then before discussing his opinion with other members he overhears someone say that this fruit has been condemned by a highly credible source, say a group of gourmet oncologists. Although the person's familiarity with the kumquat itself has not increased—it would be no easier now knowing this for him to distinguish it from another fruit—in all likelihood this additional information would greatly enlarge the universe of kumquat-relevant propositions (e.g., arguments, scenarios, and scripts about carcinogenic agents become pertinent). Certainly the ideas he

subsequently expresses in discussion will be different from those available to him initially.

The biases that concern us here stem from the fact that group members know each other's attitudes. Let us briefly consider the effects of such knowledge in relation to some well known results from Asch (1951; 1956) having to do with the size of a majority (for an analysis of other results from Asch, see Burnstein and Vinokur, 1977). A still puzzling finding is that the tendency of a group member to agree with a unanimous majority increases with the size of the majority but only up to the size of three (Asch, 1956). Hence, it would seem that a unanimous majority need not be large to exert all of its potential influence. It is important to recall, however, that in this famous experimental paradigm the attitudes of individual group members are revealed *in series*. Since the critical subject is always next to last, he receives more information about the preferences of other members and has a greater amount of time to generate arguments as a majority size increases. The knowledge that another member prefers, say, alternative *J* rather than *K*, focuses the person's attention on the former. The larger the majority, the more frequently this alternative will monopolize his attention and thereby lead him to generate a disproportionately large number of arguments explaining why *J* might be preferred. Hence, the tendency of group members to agree with a majority would continue to increase over a large range of sizes. Neither Asch nor other researchers have been able to obtain such an effect. The argument-processing analysis suggests why this might be the case. While others reveal their preferences for *J*, the person, at some point in this string of revelations, will have exhausted the pool of pro-*J* arguments. From that instant, knowing that still another member favors *J* can have no impact because the person simply will be unable to think of additional reasons for being pro-*J*. Whether this point occurs early or late in the series obviously depends on the number of pro-*J* arguments in the pool. In Asch the alternative preferred by the majority is unlikely to elicit many favorable arguments. Hence, according to persuasive-arguments theory, under these conditions the tendency to conform will not increase with the majority size beyond some relatively small value.

Keith Sentis and I are testing this analysis, and I would like to briefly describe this work in progress. We have assumed that if there is an appreciable number of arguments for *J* as well as *K* in the pool, then (1) the majority preference, *J*, biases a person's line of reasoning so that a relatively large number of pro-*J* arguments come to mind, and (2) this biasing effect increases with the time, which in the Asch situation covaries with the size of the majority. Consequently, the larger the majority, or the more time members have to think about the majority preference, the more likely it is that they will conform and prefer *J*. In our experiment, six-person groups were confronted with twenty-four issues similar to those described earlier in the study of depolarization. Members sat in separate cubicles containing a video monitor and a six-button response panel and responded in series, simulating a voting queue. An issue would appear on the monitor followed some time later by a signal to vote, the interval presumably depending on the member's position in the queue. Before indicating their position on an issue, members learned the position either of no other members (on a third of the items when they were first in the queue), of three others (on a third of the items when they were fourth in the queue), or of five others (on a third of the items when they were sixth in the queue). The pattern of others' responses displayed was either highly typical or

highly atypical for the item in question. The important variable was the length of time a group member had to consider the position of the majority. On half the items a short (10-second) interval was given, and on the other half a long (40-second) interval. The results, collapsed across the two types of items, are presented in Fig. 1. These curves describe how much the average individual preference under a three- or five-person majority (i.e., when the member was fourth or sixth, respectively, in the voting queue) deviated from that under no majority (i.e., when the member was first in the voting queue). The reader should keep in mind that when

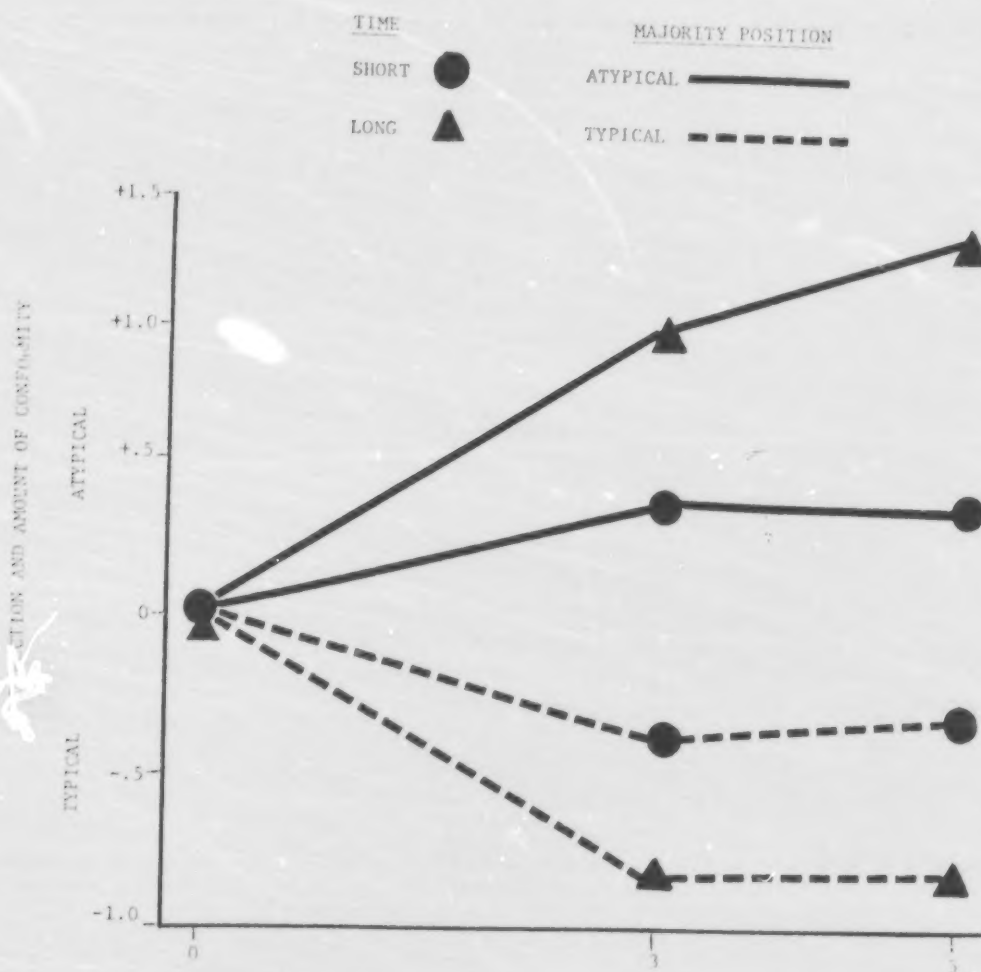


Fig. 1—Conformity as a function of time to consider the majority position when the majority varies in size and its position is typical or atypical

(Conformity is measured in terms of deviations from the position held in the absence of knowledge about the attitude of others (zero majority size). The values of the vertical axis refer to units on a six-point attitude scale.

the majority position is typical for the item in question, there is by definition little difference between the latter and the position taken by the average member in the absence of a majority. Thus, the opportunity for the majority to exert influence is extremely limited. In fact, instances in which typical distributions are presented correspond to those trials in Asch in which the confederates made the correct choice.

As expected, in the face of an atypical majority, members reliably shift in an atypical direction; and in the face of a typical majority, they shift in a typical direction, even though in the latter case the distance available for a shift to occur is meager. More interesting is the finding that the longer members have to consider the majority position, the more they conform. This holds regardless of majority size. In addition, a large majority has the same impact as a small majority *only* if members have little time to think about the majority position. When there was much time to think, conformity increased regularly with majority size, which is what an argument-processing approach would suggest.

The central idea behind our conformity research is that an opinion expressed by another member activates ideas associated with that position. Naturally, if the opinion of others remains unknown, only ideas contained in the statement of the issue and in one's prior knowledge are available for processing. Therefore, the logical complement to the first study (discussed above) would vary the time a member has to access and integrate the latter kind of information and form an opinion of his or her own *before* learning the opinion of others. In this case, a main effect would also be expected for processing time, except here the more time, the less conformity (i.e., the more time available to process one's own opinion, the less impact the others' opinions should have). We have recently carried out this study and it indeed seems to be true that a unanimous majority, large or small, has relatively little impact on group members when they have had appreciable time to form an opinion, but it has a large impact when the members have had only a short time to do so. Moreover, we found substantial evidence for an argument pool-size effect. If there are no arguments available, a member who thinks about the issue for a long time before forming an opinion will have taken into account no more arguments than a member who thinks only a short time. As a result, the former will be just as amenable to persuasion as the latter. More specifically, in our second conformity study the majority had a large influence on judgments about obscure facts, where there were virtually no arguments available, regardless of how much time members had pondered prior to expressing an opinion. In contrast, the majority influence depended markedly on time on items where there were many arguments available; that is, there was little majority influence when opinions were formed after prolonged thought and much influence when they were formed quickly.

CONCLUSIONS

These findings as well as those obtained by other researchers in the past few years suggest that we need to change our conception of persuasion, especially in group research where normative and informational processes are still seen as equally important bases for social influences. According to our analysis, the distinction between normative and informational processes is a phenotypic one. It may

remain convenient for practical purposes, but there are increasing signs that social influence can be explained primarily in information processing terms. In the area of group polarization, at least, I feel confident that the problem of social influence is reasonably well understood: It is essentially informational. Of course, settled problems are not interesting. However, I do not expect that work on this issue will be discontinued. Not only do we know very little about the nature of argument processing, but also, many of my colleagues are probably sure this analysis is incorrect and will act accordingly.

The second point I want to make is a programmatic *caveat*. To say as we did at the very beginning that the small group is an information-processing system implies that the process by which it reaches a decision can be decomposed and that its capacity for operating on information is limited and thus selective. However, our research is typical of the work on small group decisionmaking in at least one respect, in that it confines itself to one stage. In this stage, the most feasible options have been already laid out, fixed by the experimenter, and there is only a single selection (information limiting) routine appropriate, namely, choosing the best option. To understand the small group in a broad range of settings, however, we must study it as it moves through a "natural" performance sequence, for example, the series of stages inherent in group problem solving. Because time is so limited I will try merely to give you an overview of this sequence, one that we have used in other research not discussed here (see Fig. 2). Even a cursory examination of this descriptive model raises several basic issues on team performance, particularly in respect to the temporal organization of problem-solving and decisionmaking operations. For example, how does the group *recognize* and *diagnose* a problem? How does it *search* "group memory" for solutions and then *screen* these for feasibility (or design one when there is nothing feasible)? How does it *evaluate* and *choose* among more than one feasible solution, etc.? Finally, and most importantly, under what conditions does the group *cycle back* (perhaps recursively) through any one or more of these stages? Neither research on group polarization, nor on social decision schemes, nor on any of the other popular problems in the area of small groups comes close to answering such questions. However, I think these are the issues with which we will have to grapple eventually. This conference, with luck, may provide the opportunity to begin.

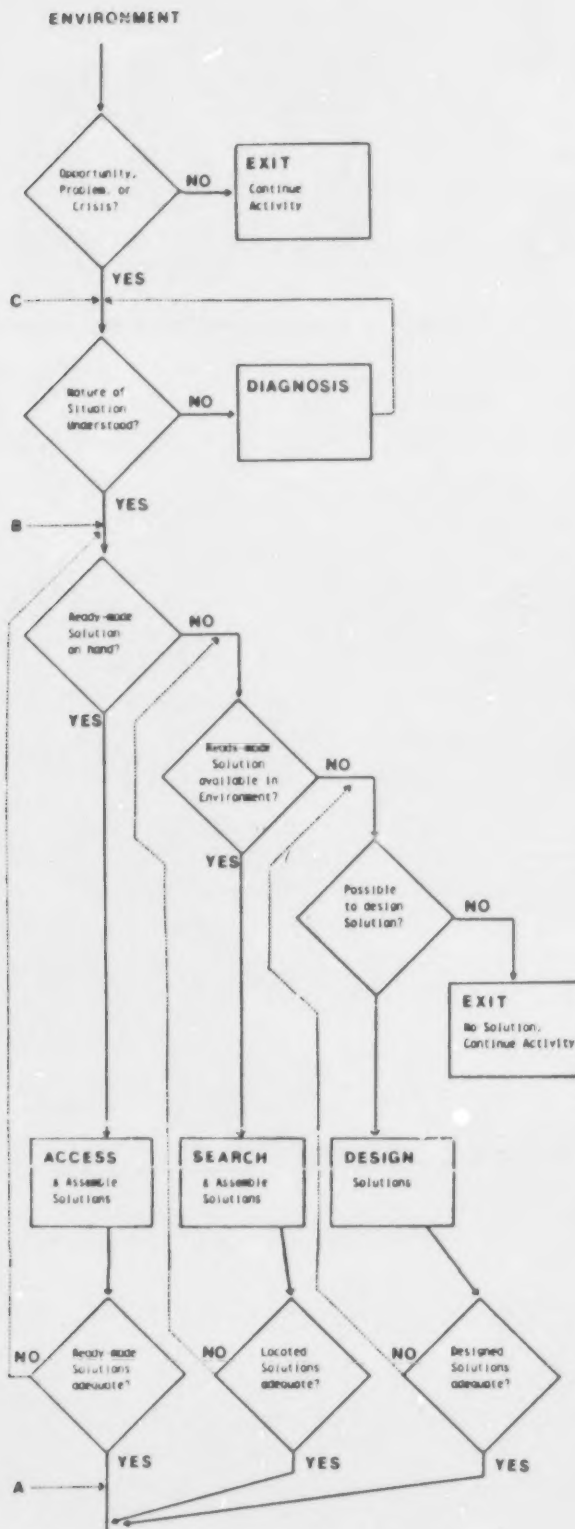


Fig. 2—Stages in group decisionmaking

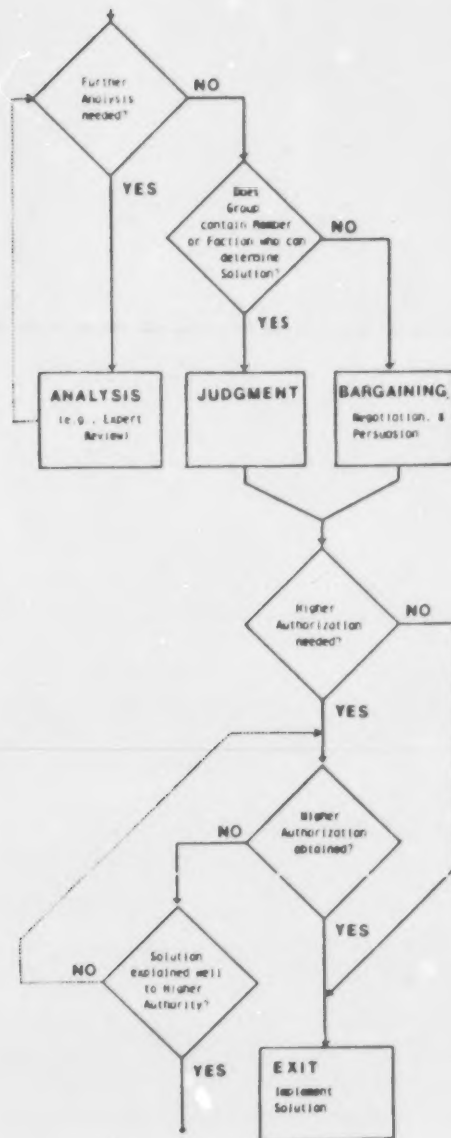


Fig. 2 (Continued)

SELECTION

INTERRUPT TO RECYCLE

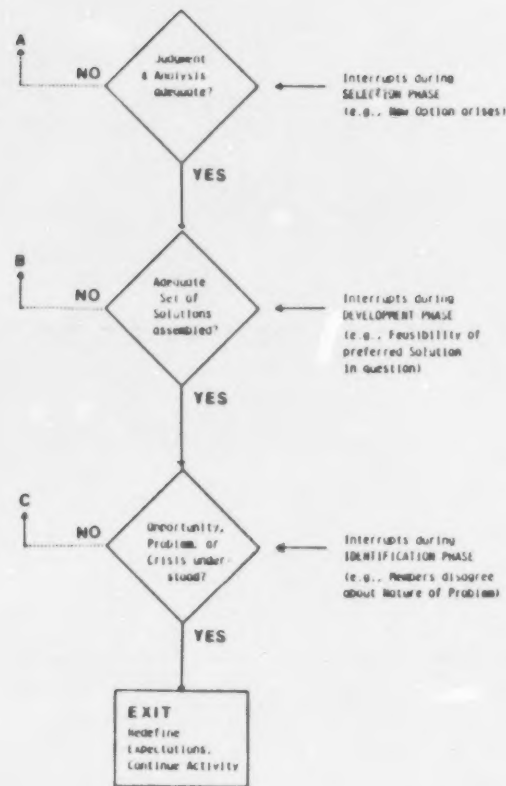


Fig. 2 (Continued)

DISCUSSION OF SESSION 3

Thorndyke began the discussion by asking how Burnstein's methodology might be applied to improving tactical decisionmaking. Burnstein replied that the arguments used in his experiments could be cast in an *if-then* format typical of tactical decision situations. He also noted that the processes observed in the laboratory had also been studied in such real-world situations as Israeli military decisionmaking and decisionmaking within medical teams. Most importantly, in his opinion, his research demonstrated that the semantic content and information value of individual arguments are crucial in determining the final decision. However, he acknowledged that his work addressed only a single phase of the decisionmaking process. It did not consider how teams recognize or diagnose a problem, or how they generate potential solutions, but only how they select among alternative pre-established solutions. Other procedures, such as Bales' system for analyzing interaction, could be used to study these earlier stages in the decision sequence.

Barbara Hayes-Roth commented that Burnstein's work made two significant points. First, people do use the information provided by other group members in an intelligent way when making a decision. Second, certain arbitrary factors, such as the time spent thinking about an argument or the order in which arguments are presented, also influence decisionmaking. She suggested developing protocols for group interaction that would control non-semantic biases and thus provide pure estimates of the effects of information content. Several participants pointed out examples of such techniques, including the Delphi procedure and military court martials, in which officers must present their arguments in reverse order of rank.

Shifting to the topic of social factors involved in persuasion, Burnstein noted that in the study of medical teams mentioned earlier, he had found evidence for self-censoring caused by power relations. That is, individuals with the lowest status will not offer all of their arguments. He and his colleagues had expected to find that the doctor in the team held the highest status, and the social worker, the lowest. In reality, these positions were reversed. The statements of the social worker carried the greatest weight because she interacted daily with the patients and knew the most about each case. This example illustrates the distinction between formal and informal group structure. The institutional hierarchy yielded to a more pragmatic, knowledge-based hierarchy.

Nilsson commented that "group culture" may partially determine the behavior of decisionmaking groups. Some groups, for example, will have a tradition of decisionmaking by consensus, while others may anticipate continuing factionalism and dissent. Burnstein suggested that such historical information about the group could be included in his model as part of the "knowledge reservoir" from which arguments are drawn.

Irwin Goldstein commented that data on decision outcomes could be misleading without process information. He maintained that measures of the decision process are needed to guide interpretation of final decisions.

When asked to suggest further applications of small group research to team issues, Burnstein cited sociobiological factors in group interaction as one promising area. For example, sexually, racially, or ethnically mixed teams may experience

unique team problems, especially when the team lives and socializes together for long periods of time. King noted that some aircraft carriers had been unable to operate due to racial tension.

Weiner suggested that Burnstein's results would not be affected by sex or race differences. However, Burnstein maintained that if discussion centered on an appropriate issue, sex or race differences could significantly influence group decision-making. Barbara Hayes-Roth summarized research showing that even in discussions on a neutral issue, women contribute less than men when the group interacts face-to-face. This difference disappears, however, when group members communicate via computer-linked terminals.

Tolcott asked how one might induce groups to generate a rich set of alternative strategies for action. He suggested that since group heterogeneity influences the selection of an outcome, individuals with distinctive knowledge and backgrounds might generate a broad set of outcomes.

V. COGNITIVE PSYCHOLOGY

A COGNITIVE SCIENCE APPROACH TO RESEARCH ON TEAMS

Allan Collins

Cognitive science attempts to analyze human performance in terms of the formalisms of artificial intelligence, linguistics, and philosophy. By considering three potential research areas, I would like to show how a cognitive science analysis could produce valuable research on team performance and training. It is possible to analyze naturally occurring situations through task analysis and dialogue analysis, in order to identify the important variables. Then we can construct laboratory experiments to test the effects of different process and structural variables. Both analytical and experimental approaches will be suggested for dealing with the three kinds of issues I will raise.

GOAL ANALYSIS OF TEAM SITUATIONS

Central to most computational analyses of human performance is an analysis of the goals and subgoals driving the behavior. This approach started with the means-ends analysis of Newell and Simon (1963), but carries through the literature on problem solving and planning (e.g., Newell and Simon, 1972; Sacerdoti, 1975; Schank and Abelson, 1977; Brown, Collins, and Harris, 1978; Bruce and Newman, 1978; Hayes-Roth and Hayes-Roth, 1979). Goal analysis is particularly central to team performance because the goals of the different members can interact in a variety of ways.

I can illustrate how important goal analysis may be to team performance with an anecdote from Oliver Selfridge. Suppose a mother wants her yard to appear neat and clean to her neighbors. The father might avoid raking the yard himself by offering his kids a dollar for each large bag of leaves they collect. The kids, who are interested in making money, might maximize their profits by collecting leaves from the neighbors' yards as well as their own. What has happened, and what frequently happens in multiperson situations, is that the goals of the team members lead to consequences not intended by the team leaders.

In goal analysis, applied to team performance, the pursuit of goals and subgoals by any member must be analyzed in terms of how the intended effects and side effects influence the attainment of the others' goals and the group's defined goals. In particular, there are at least four kinds of interactions that might occur between different members' goals:

1. Realization of one person's goals is *identical* to achieving another person's goals or the group's goals.
2. Realization of one person's goals *facilitates* achievement of another's goals or the group's goals.
3. Realization of the person's goal *interferes* with achievement of another's goals or the group's goals.

4. Realization of one person's goals is *irrelevant* to achieving another person's goals or the group's goals.

Both facilitation and interference occur by satisfying or disabling preconditions necessary to achieve others' goals.

The kinds of hypotheses that might be tested in this regard involve how different aspects of group structure (hierarchical structure, flexibility, size) affect the coordination of the group members' goals. For example, it may be the case that rigid hierarchical structures are particularly prone to mismatches in goals between members at different levels. However, flat structure may only be possible when all the members of a group are pursuing identical goals. It may also be that stable subgroup structures spawn autonomous subgroup goals that do not coincide with the overall group goals.

These questions can be investigated effectively both by task analysis of team performance and by structuring groups in different ways to carry out specific tasks. Task analysis should be carried out on different groups with a variety of both missions and structures. A task analysis should involve systematic interrogation of group members as to their explicit goals and subgoals, and observational analysis of how their actions affect other members' attainment of goals. Experimental studies can be carried out to vary different aspects of groups (the command structure, role flexibility, size) to test specific hypotheses about team interaction, such as the proposed hypotheses about hierarchical structures above.

COMMUNICATION ANALYSIS OF TEAMS

There are two areas I see as potentially valuable to communication analysis of team performance, which I will call *communication failure analysis* and *continuation analysis*.

Communication Failure Analysis

Starting in the philosophical and linguistic literature (Austin, 1962; Searle, 1970; Cole and Morgan, 1975) and spreading more recently to the artificial intelligence and psychology literature (Clark, 1979; Cohen and Perrault, 1979), there has been extensive analysis of different kinds of speech acts that occur in discourse and of the preconditions (called "felicity conditions" by Searle) that must be satisfied in performing any speech act. This kind of analysis has been extended to sequences of dialogues by Levin and Moore (1977) in their work on dialogue games. In particular, Levin and Moore specify a number of different types of utterance that frequently occur in dialogues (requests for help, requests for information, etc.) and the structures and preconditions the participants assume in carrying on different types of conversation.

In analyzing team interaction, we may find that miscommunication arises when there are mismatches between the assumptions about communication structures by different team members. When a team remains together for a long time, such mismatches would tend to dissipate, but teams with high turnover may suffer from severe failures of communication. One of the benefits of rigid role structure may be to minimize communication failure when there is high turnover. Groups linked

together by electronic means (either voice or mail) may present many problems of miscommunication.

These ideas might be tested initially by analysis of the conversational interactions of groups with different structures and different turnover rates. This work can be based either on analysis of transcripts or on behavioral observation. Any analysis should try to identify whatever mismatches occur in communication between different members. These preliminary investigations should lead to experiments to test specific hypotheses about how different variables affect communication between team members.

Continuation Analysis

In recent years I have been analyzing teaching dialogues to determine the kinds of situations that elicit different questions and comments from teachers (A. Collins, 1977; Stevens and Collins, 1977; Collins and Stevens, in press). This kind of approach has promise for analyzing group interaction, particularly in problem-solving situations such as crisis management.

In particular, it is possible to classify different types of continuations (i.e., relations between an utterance and a preceding comment) in terms of their efficiency in leading to the group's solution of a problem or completion of a task. I would suggest the following initial taxonomy of continuations:

- *Extrapolation*: Taking an idea and showing how it can be extended, applied in a new situation.
- *Critique*: Pointing out the holes or limitations of a previous idea.
- *Integration*: Showing how an idea is related to previous ideas.
- *Reformulation*: Explaining an idea in clearer, more precise terms.
- *Extrapolation question*: A question asking if a person's idea could be applied in a different situation.
- *Clarification question*: A question designed to elicit reformulation by the previous speaker.
- *Critical question*: A question asking about possible limitations or holes in a previous speaker's ideas.
- *Integration question*: A question asking how a person's idea relates to other ideas.

My hypothesis is that certain types of continuations are more profitable in different tasks. For example, in brainstorming, the most profitable continuations may be extrapolations and integrations, whereas in crisis management, critiques may be quite important. Thus the most interesting issues to study are the interactions of task variables with frequency of different continuation types.

Both training and selection of people for teams might be based on continuation types. Some people are particularly good critics; others are particularly good at generating novel, though perhaps ill-formed, ideas. If particular continuation types are most profitable for specific tasks, team members might be given training in how to generate comments of a particular type.

A COGNITIVE SCIENCE AND PSYCHOMETRIC APPROACH TO TEAM PERFORMANCE¹

Earl Hunt

In this paper I shall discuss some of the recent developments in both cognitive science and psychometrics and then apply these ideas to the evaluation of team performance. Cognitive science has grown out of various attempts to specify the computations that might define individual thinking. I shall try to show that cognitive science studies of individual thought also suggest ways of thinking about groups. The question of evaluation is more difficult. One of the chief criticisms leveled against cognitive science by more traditional psychologists is that cognitive science enthusiasts pay very little attention to evaluating the ideas that they develop. This has proven something of a problem in my own work, for I have been interested in applying the ideas of cognitive science to gain a better picture, not of thinking in general, but of changes in thinking both between and within individuals. Over the years we have gathered a very large body of data about such changes, as witnessed by the literature on intelligence and on group and sex differences (Willerman, 1979). It is not enough to have bright ideas about changes in thinking; bright ideas must be matched against this literature. Fortunately for us, the evaluation issue has been treated in great detail within modern psychometrics, and we have been able to apply some psychometric evaluation methods to examine ideas that were derived from cognitive sciences.

The fact that cognitive science ideas have been evaluated within the individual differences field is, in itself, only of passing interest to those who are interested in teams, unless some analogy can be found between the analysis of individual thought and the analysis of team performance. I shall argue that there is such an analogy. To develop this argument I shall first describe some of the ideas that cognitive science has developed about information processing systems *in general*. The "in general" is important, since many team performance situations can be thought of as an exercise in information processing. I will then make a detour to consider the analysis of covariance structures, a psychometric technique for data evaluation that has proven useful in my own work on the evaluation of cognitive science models. Finally, I shall try to draw these threads together by outlining a possible application of cognitive science models plus psychometric methods to the study of the interplay between leadership and individual performance in team situations.

COGNITIVE SCIENCE MODELS OF INFORMATION PROCESSING

The cognitive sciences deal with the flow of information in problem solving by studying such disparate problem-solving situations as chess and medicine. They

¹ This research was supported by the Office of Naval Research through a contract (No. N00014-77-C-0225) to the University of Washington, Earl Hunt, Principal Investigator.

provide what could be called an empiricist approach. There is an assumption that problems and their solutions can be represented by building complex data structures from a small set of "basically perceivable" primitive elements. The prototypical cognitive science approach to understanding an information processing system (usually, but not necessarily, a human working on a problem) is to design a computer program that would do the same thing. What we get from cognitive science is an analysis of the way things "have to be," a sort of commentary from a philosopher with an engineering degree. This can be very useful. It can also be fraught with fads, as new programming techniques are mistaken for basic truths about the organization of knowledge. Fads have certainly produced many detours in the field. In spite of this, there has been an important evolution in our thinking about the organization of problem-solving systems.

The initial efforts in the field, sparked by Newell and Simon's (1972) work on the General Problem Solver (GPS) program, were aimed at designing a "universal" problem-solving system in which specific problem-solving techniques were to be treated as special cases.

The GPS program can be thought of as consisting of two parts: operators and an executive. The operators were the part of the program that actually did the work, i.e., only operators could change the internal representation of a problem. The job of the executive was to choose which operators to apply, and in what order. To do this, the program established a hierarchical system of goals and subgoals. The flavor of the system can be seen by examining its flow diagram, which is shown in Fig. 1. This depicts a highly recursive structure. To GPS, a problem consisted of two descriptions, a "present state" and a "desired state." The executive would compare the two and determine the difference between them. It would then determine what operator was appropriate for reducing the major difference between the two descriptions. If the operator could be applied to the present state, it would be. If not, the program would set up the subproblem of changing the present description so that the desired operator could be applied. This problem would then be solved, recursively, as a GPS problem. Naturally, attacks on a problem might lead to the generation of sub-subproblems, etc.

I do not want to get into the programming details of how GPS did all these things. The important point is to note the philosophy behind the program. All actual operations on the data were to be tightly under the control of the central executive program that selected the operators. At any point the executive program "knew" why it was applying an operator, in the sense that each operator was applied with a particular goal in mind. This meant that the executive had to have a good deal of information about what the operators could do. Goals and subgoals were pursued in a rather single-minded manner; the program worked on one path until it was blocked, and then it took up another approach. Most important, this problem-solving method made it difficult (although not impossible) to pass information from one operator to another, except at the specific moments at which the executive directed that information be passed.

The GPS is a good model for some sorts of problem solving. Planning a specific trip is done this way. To get from the University of Washington to The Rand Corporation, you first must fly from Seattle to Los Angeles; to fly you must have a ticket and you must be at the airport; etc., etc. Intuitively, it might also serve as a model of tightly organized team performance in some situations. One could think

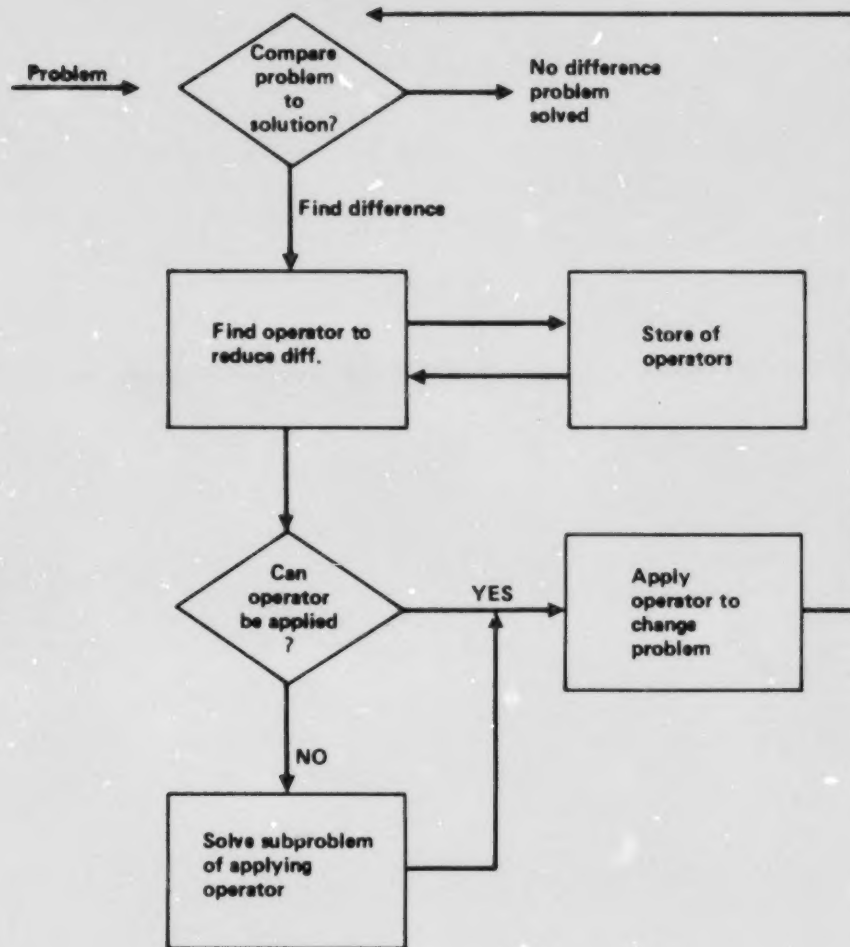


Fig. 1—GPS ("centralized") organization

of evaluating a football team in the same way; by looking at two things, the efficiency of local operations and the efficiency of the selection procedures that determine the sequence of operations to be applied.

As cognitive science studies progressed, and in particular, as attempts were made to study less structured situations, it became clear that this model was rather limited. It is not always possible for a general executive to lay out an appropriate structure of goals and subgoals, independent of any knowledge about the problem domain itself. Also, and especially in situations involving "expert knowledge," it appeared that we had to allow for communication between the operators independent of the functioning of the executive. A rather different problem-solving organization has evolved in response to these challenges. In this system much more of the burden for operator selection is placed upon the operators themselves. The program organization is depicted in Fig. 2. It consists of an "active memory" section

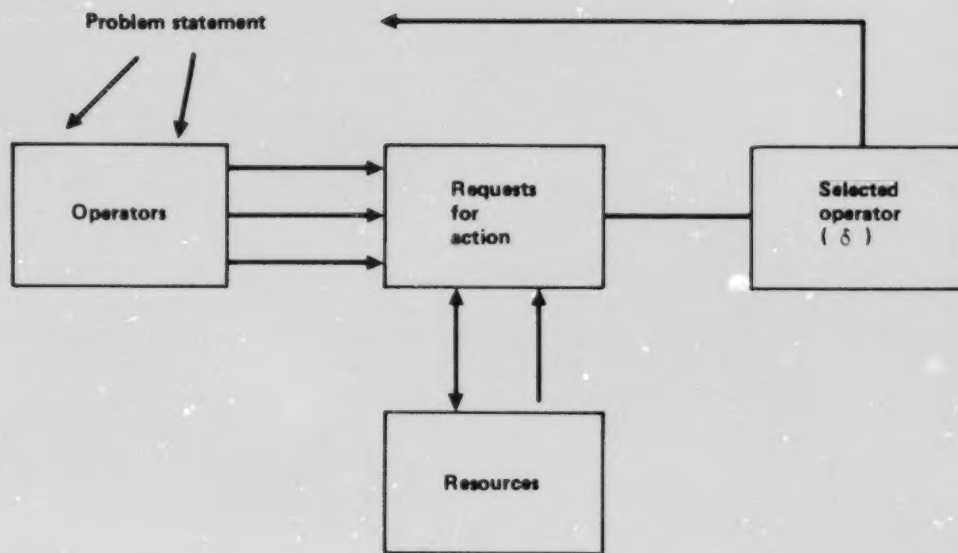


Fig. 2—A decentralized problem solving organization

that holds a limited amount of relevant information about the current problem, including requests for action, and a "long-term memory" system that holds a set of rules (the operators) about how to do things. Each rule consists of a *pattern recognition* and an *execution* part. The pattern recognition part looks at the current state of active memory and recognizes situations in which its actions "might help." When such a situation is recognized, the rule requests permission to execute its action. To gain some appreciation for this, consider the familiar simplification rule of algebra:

$$CA + CB = C(A + B).$$

This can be looked upon as a rule that says, "Whenever you see a structure of form $CA + CB$, request permission to rewrite it in the form $C(A + B)$."

The job of the executive in this sort of system is to react to the request for execution and to resolve conflict when two or more rules recommend incompatible actions. A great deal of local knowledge can be built into this system by designing complex operators, without requiring that the executive be made similarly problem-specific. In fact, in the case of ultimate generality, the executive could be a simple system for selecting the strongest of n signals that varied over time, without any regard to what those signals were (Hunt, 1980). A good deal of communication can take place between operators by having one operator plant information into active memory for later use by another operator. This system will work even if the operators involved are not executed one after the other. Indeed, one operator may not be primed to fire until each of several preceding operators, each working on a different aspect of the problem, independently deposits bits and pieces of the information that will eventually trigger the final operator. The executive will have

relatively little control over such an interchange and hence does not have to have the problem-specific knowledge that would be required to control it.

These models are end points on a continuum, rather than mutually exclusive alternatives. (Indeed, it is even possible to program the first by using an interpreter built on the principles of the second method (Newell and Simon, 1972).) Although the continuum arose from the system of problem solving "inside the head," there is no reason that it cannot be applied to describe the organization of problem-solving teams. Cognitive science studies have indicated (*not proven*, for cognitive science remains weak on evaluation) that centralized problem solving works well when there is a clearly defined problem with the solution and its restrictions easily recognizable, and when the operators available are limited in number and fairly easy for the executive to understand. Decentralized organization seems to be more appropriate for ill-defined problems where the ultimate goals are not specifically stated, and where there are a very large number of options. To illustrate in a team context, and using some very complex examples, the 1979 Iranian hostage crisis seems to me to be a situation of the first type. The goal—return of the hostages without appearing to yield to blackmail—was clear, the options were limited, and their probable effects statable with some precision. The "war against inflation" is a situation of the second type. There are multitudinous goals, and it is hard to define the intermediate steps toward a final solution. The only problem-solving method feasible seems to be to orchestrate local solutions in the hopes that as different operators are applied—minimizing conflict between them—things will gradually get better. I have noticed with more than passing interest that the President's office assumed the role of a GPS executive in the first crisis but plays the more passive role of reacting to local alternatives in the second situation. From a cognitive sciences perspective, this choice of roles was as much forced upon the executive by the nature of the two problems as it was made because of the nature of the actors involved in 1979 statesmanship.

Suppose you accept my characterization of problem-solving methods and even are willing to allow that it can be applied to the description of team problem solving. What about evaluation? The next section treats this problem as a completely technical one in measurement and statistics. Only after the technical details have been presented will I deal with the substance of evaluation. While reading the technical details, though, it is worth keeping one point in mind: Consider the division of team members into "management" and "operators." (Of course, a person might simultaneously be management and operator, as is the pilot of an airplane.) The importance of individual skills to team performance will depend upon the problem-solving organization being used and upon the nature of the problem being solved. Most particularly, the "skill" of knowing *when* to take a particular action must reside in management in the first organization and in the operators in the second. Problem-specific knowledge (or the ability to develop it) is required by management in a centralized organization. In a decentralized organization, management, strictly speaking, never has to know what is going on—all the manager has to be able to do is to recognize who does know and to give them the resources they need. In the final section I shall try to connect these prosaic observations to the technical material to be given below in an attempt to analyze an actual problem in organizational behavior.

A PSYCHOMETRIC TECHNIQUE FOR IDENTIFYING AND EVALUATING LATENT TRAITS

In this section I will briefly describe a psychometric method known as the analysis of covariance structures (ANCOS). For those who are interested, it can be looked on as "On beyond factor analysis." The technique itself was developed by Joreskog (1974). Discussions of its application can be found in original research by Bentler and Speckart (1979) and in a review by Bentler (1980). I personally became aware of it when we were faced with problems in analyzing the role of visual information processing in reading (Palmer et al., 1979), and it has subsequently figured largely in my thinking about the testing of complex models of how people do two things at once. What follows is a strictly non-mathematical presentation of the logic.

The method is appropriate for situations in which we are interested in understanding the relationship between several imperfect measures of "latent traits." Most measures of team performance are of this nature—we seldom have one observable measure that we are willing to say is *the* measure of group performance. Instead we are usually faced with a multitude of measures, each of which has some imperfect claim to being a measure of some aspect of team functioning. Suppose that we have taken multiple performance measures on each of several teams. We want to ask whether or not it is reasonable to order these teams along a single dimension, from "good" to "bad." We can answer this question by analyzing the correlation matrix *between* the measures of individual team performance. Figure 3 shows the logic of the approach by illustrating the possible measures. Each of the n measures is assumed (a) to measure some specific factor that is unique to that measurement and (b) also to have a correlation, f (technically called a *loading*), with the hypothesized dimension of good and bad team performance. If this is the case, then the *correlation* between two measures of team performance should (within some margin of statistical variability) be equal to the product of the loadings of each measure on the single dimension of good and bad performance. That is, for any two measures M_i and M_j , we require that:

$$r_{ij} = f_i f_j$$

The ANCOS procedure is simply a mathematical technique for finding whether or not appropriate f loading factors, $\{f_i\}$, can be found. If they can, then each team can be assigned to a point along the "good and bad performance" dimension.

It may have occurred to the reader that what I have done is to give a very complicated description of a factor-analytic technique for solving for a single general factor. That is true, but I have only used this degenerate case to set the stage. Assuming that teams are, generally, at least as complex as the people in them, most investigations of team performance will discover that the simple model of Fig. 3 is inadequate to account for the data. To give the discussion some context, suppose that we assume that a team can be characterized by two "traits," task effectiveness and morale; and that the two, although correlated, are not identical. Further, different measures may be looked upon as reflecting one or both traits. This sort of structure is shown in Fig. 4. Note that it subsumes the structure of Fig. 3, since Fig. 4 would become Fig. 3 if the correlation between the two traits were one. Thus any data that can be fit to Fig. 3 can be fit to Fig. 4, but the converse is not true.

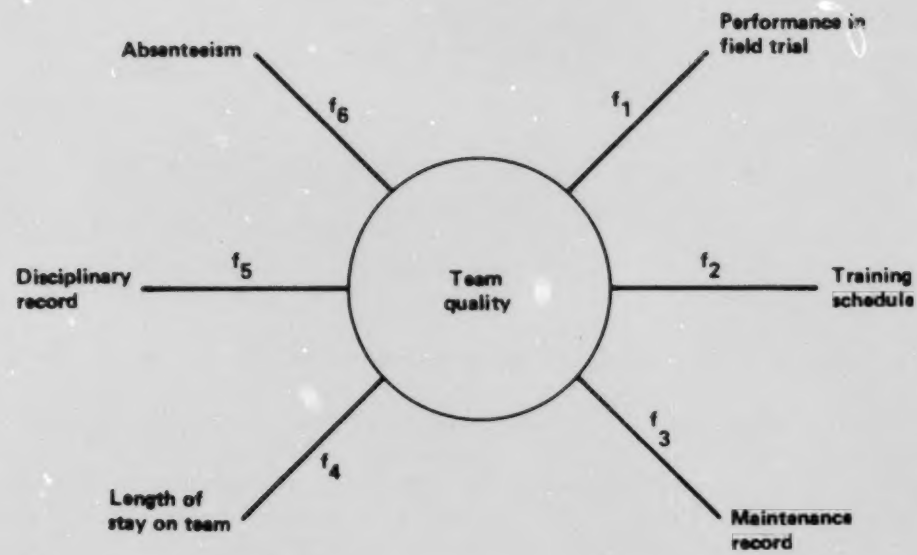


Fig. 3—A very simple model of unidimensional team performance

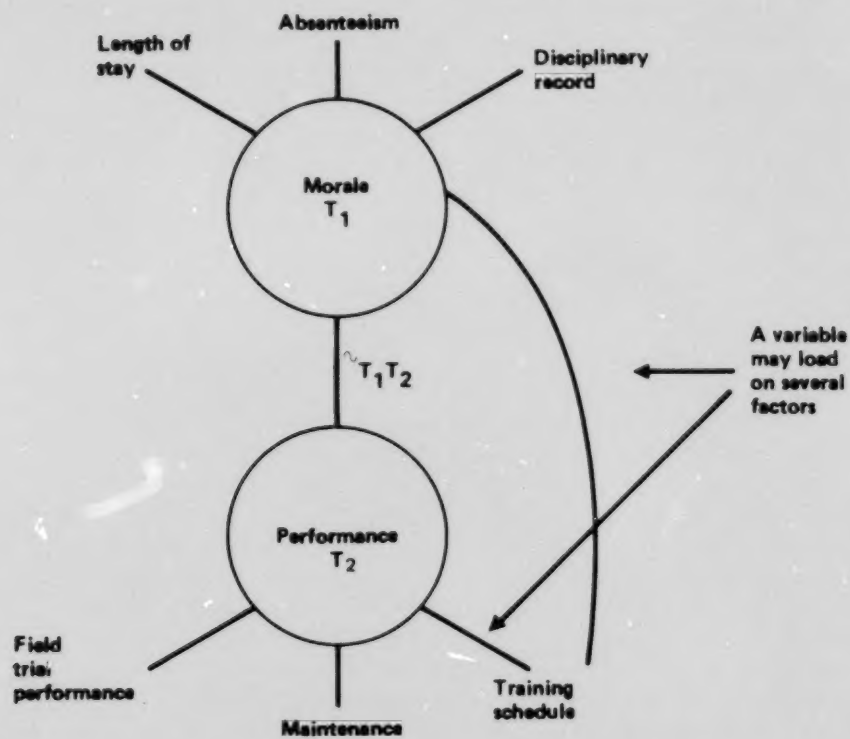


Fig. 4—A more realistic model

I will simply state that there are statistical techniques for evaluating the relative fit of models in such a case, without going into any detail about what they are.

Now, let us examine a more interesting model which will bring us back toward cognitive science. We might suppose that, in general, one could characterize individuals by (a) their possession of job-relevant skills and (b) their possession of interpersonal skills. We assume that these can be measured, though imperfectly. We further specify that people may be either "management" or "operators." We could then test a complex model which says, roughly, that team performance is characterized directly by performance and morale, and that these depend both on management and task skills, filtered through the organization of the team. Figure 5 shows a model in which we postulate that individuals have both task-relevant and interpersonal characteristics, and the leader has task-relevant and leadership skills. These, in turn, have varying direct or indirect effects on team morale and effectiveness. The model can be tested by examining the covariance matrix that combines both task-relevant and interpersonal characteristics of team members (measures $\{M_i\}$), similar characteristics of the leadership (measures $\{L_j\}$), and various measures of team performance ($\{O_k\}$).

In order to bring our cognitive science models into the picture, we need to consider just one further class of measurements. These are the measurements that tell us what sort of system we are dealing with—a centralized or a decentralized problem-solving system. The theory defines the sort of measures that we should have. The two types of problem-solving systems differ along two dimensions: the extent to which the leader is responsible for setting goals and subgoals, and the extent to which an individual worker is responsible for *initiating* a proposal that a particular action be taken. (In both systems the leader decides what action will be taken.) While these behaviors will typically be negatively correlated, one is not the converse of the other. Presumably, behavioral measures could be taken of either of these dimensions. The model also forces us to make a further distinction between individual skills relevant to doing something and knowing when to do it.

The resulting "complete" model of task performance is shown in Fig. 6. Roughly, it states a model for covariance structures that makes the following assumptions:

1. Team output measures jointly reflect morale and effectiveness, and the two are correlated. Different measures are weighted toward either morale or effectiveness.
2. Team effectiveness is determined by effectiveness of planning, plus individual skills of the team member.
3. Group cohesiveness and plan effectiveness contribute to morale.
4. Planning is determined jointly by centralized planning influences via the leader and membership planning (decentrality of decisions).
5. Leader planning is determined by the leader's task-relevant planning skills, plus organizational support for a centralized structure.
6. Member planning is determined by member planning skills, plus leader coordinating skills, plus the organizational structure supporting member initiatives.

My point is not to argue for this model in any detail. It is intended as an example to show two points. We can construct a fairly complex model of the interplay

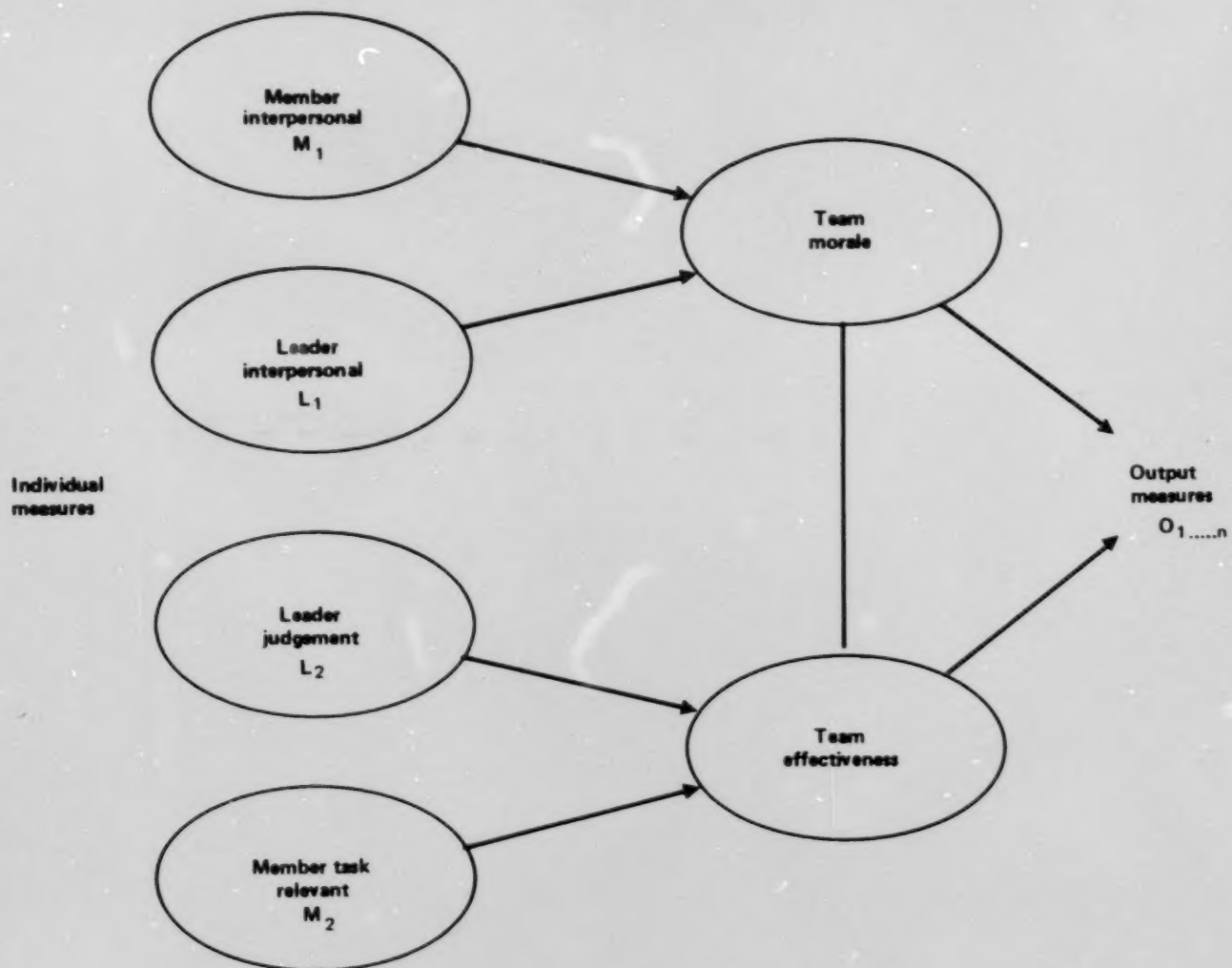


Fig. 5—Differential contributions of variables to aspects of performance

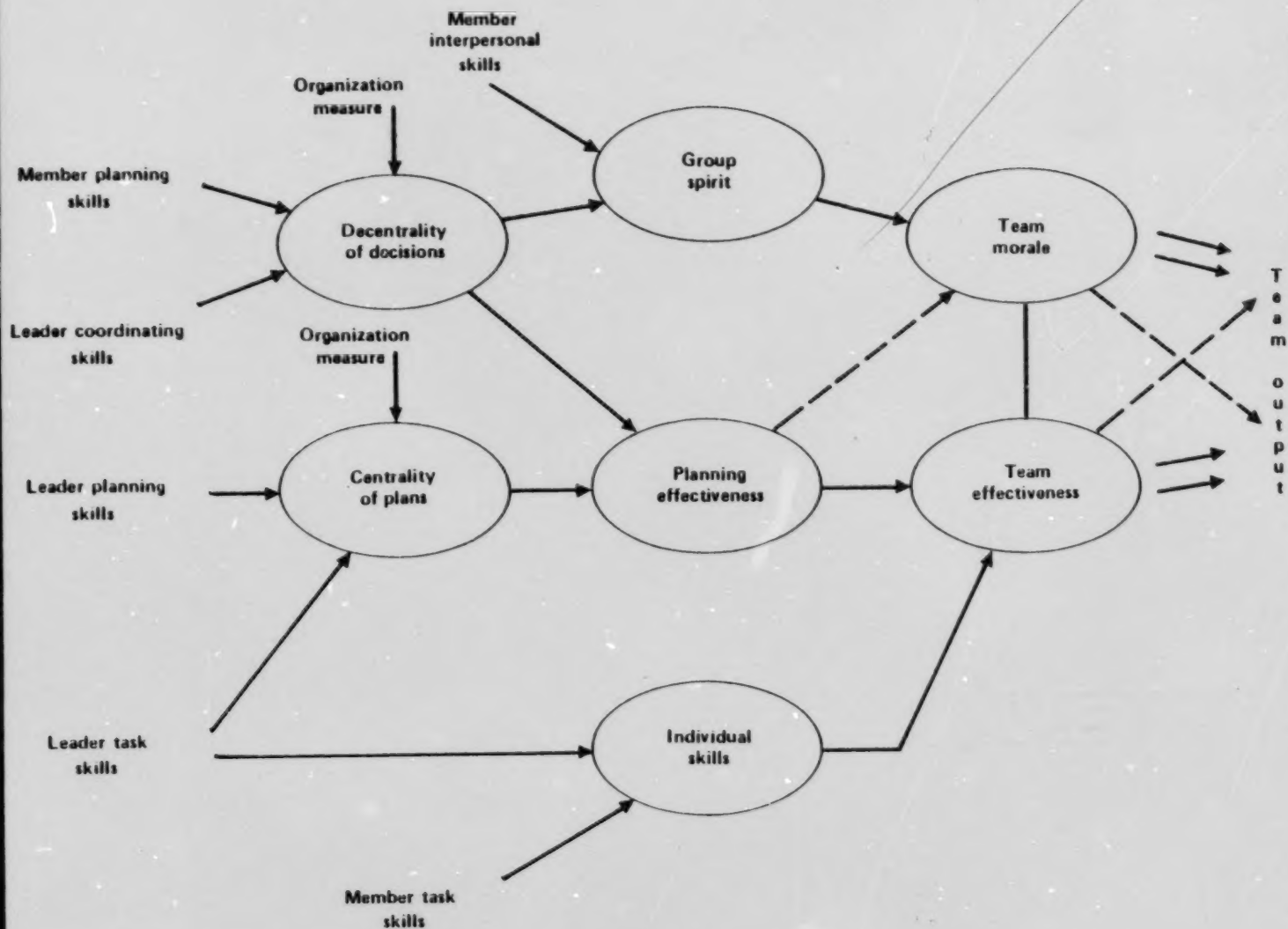


Fig.6—A “cognitive science” model of team effects

between measures of individual and team performance and organizational structure, by considering the cognitive science alternatives to the design of problem-solving systems. These models may, in turn, be tested by complicated regression analyses, using path analytic techniques based on the analysis of covariance structures. A path-analysis evaluation is far more practical than it was a few years ago, because of the development of efficient computer programs.

A FURTHER SPECULATIVE EXAMPLE

To this point, the discussion has been of abstract tasks and measures. As an illustration, I would like to outline how the analysis described here might be applied to explore some observations on organizational behavior and social psychology. The example that I will use is based on Fiedler's (Fiedler, Chemers, and Mahar, 1976) analysis of the roles of personality and intelligence in group performance. This work has generated a good deal of controversy, most of it directed toward the particular measures and designs that Fiedler has used in his research. I shall ignore this. Rather, I want to consider the basic theory and how it could be fit into the framework I propose here.

Fiedler argues that there are essentially two styles of leadership. One is "person-oriented"—a leader with this sort of orientation (a "high LPC leader" in Fiedler's terminology) devotes his or her energies to facilitating intragroup processes. Opposed to this is the "task-oriented" (or "low LPC") leader who is supposed to devote primary energies toward task accomplishments, spending relatively little time on interpersonal relationships. Fiedler regards these styles as more or less fixed traits of adults.

Instead of arguing that one or the other style of leadership is good or bad, Fiedler has put forward his "Contingency Model" of leadership, which says that each style has its place. If the situation is either extremely favorable or extremely unfavorable to the leader, task-oriented "low LPC" leaders are said to be most effective. On the other hand, in situations of intermediate favorableness, the person-oriented style of leadership is most favored. More recently, Bons and Fiedler (1976) and Fiedler et. al. (1979) have extended the contingency model in an interesting way. Based upon observations of Army infantry units, they concluded that in situations in which task-oriented leadership would be appropriate, *leader* intelligence proved to be the most effective predictor of group performance. In situations in which person-oriented leadership was important, the mean intelligence of *group members* proved to be the best predictor of team performance.

While I (and others) have been impressed with Fiedler's analysis, his explanations seem to suffer from a serious conceptual weakness. The definition of a situation as being favorable or unfavorable toward a leader is at best a slippery one. The definitions used in the many studies that Fiedler has reported, while not inconsistent with each other, have often been based on highly varied criteria. What I would like to suggest—as an alternative hypothesis to be explored—is that the key variable is not situational favorableness, but rather group structure and task demands. Further, I think that these variables can be incorporated within the "cognitive science" derived model presented in Fig. 6. To see this, consider some extreme cases. First, suppose that a group is organized in such a way that there is no possibility for centralized planning: the only thing the leader can do is to respond

to suggestions made by individual members. (This is a model approximated by all too many faculties; but the professors like it that way . . . until they become deans.) This leads to a degenerate case of Fig. 6 in which centralized planning is, in effect, a dead end. This is shown in Fig. 7(a). The leader's leverage in this case is through his or her influence on membership planning, i.e., by selecting plans proposed by the membership rather than by producing centralized plans. This will put a premium on coordinating skills. By contrast, we can consider the case of an organization that is designed to be run from the top down, in proper General Problem Solver style (with accent on "General," as in General or Admiral): In this situation, leader coordinating skills are not needed, but personal task-relevant skills, such as (usually) intelligence, are required.

To make this argument complete, we have to show that it encompasses Fiedler's findings concerning leadership style and situational difficulty. (In spite of there being some problem with many of the individual studies showing this relationship, I am impressed by the weight of the evidence.) I suggest the following argument. First, if leaders have a great deal of organizational power (e.g., division commanders rather than university deans), they can and, given our mores, probably will, enforce the centralized planning style on an organization. This is quite effective if the problem before the team is straightforward, since the centralized plan will be the appropriate one. If the problem is hard, there will be a great deal of variance in team performance, but that variance will be controlled in large part by the leader's task-relevant skills, since they will be what determines the plan. Now consider the case in which leaders have relatively low power. A centralized planning style can still be adopted for either an easy problem, in which case the plans produced by individual members will not conflict with the central plan, or for a difficult problem, in which case no one has any idea what to do, so there are no individual plans that are put forward with any great confidence.

The situation is rather different when problems are of intermediate difficulty and when the team has opted for a decentralized planning style either because of organizational strictures or personal preference. In both these cases, members will produce proffered solutions to problems and subproblems, and leaders will exert their influence by their ability to coordinate proffered local solutions rather than by their ability to produce candidate solutions themselves.

What has just been said is (intentionally) a loose restatement of Fiedler's Contingency Model in the terminology used in this paper. We can use the cognitive science approach to go beyond the characterization of situations as "favorable" or "unfavorable." The message that I, at least, draw from cognitive science studies of problem solving is that the optimal organization of the problem solver varies as a function of the nature of the problem. When solving a problem requires the execution of a fairly long chain of reasoning, but in each step in the chain there are only a few action alternatives, then centralized problem solving seems best. When the problem to be solved has many aspects, long-range planning is difficult. Thus, when there are many sources of information to be considered at each step, it seems that the decentralized mode of problem solving is preferred. This suggests that in assembling teams for problem solving we should consider a two-step operation. Based upon an analysis of the problem requirement, we choose the type of team organization required. Once this has been established, we can consider the types of individuals needed to fill the slots on the team.

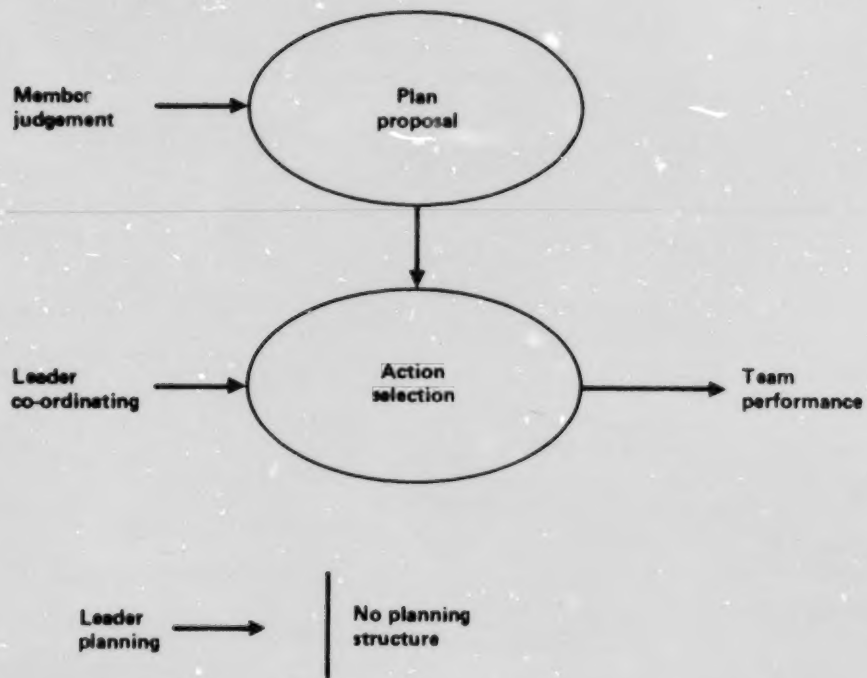


Fig. 7(a)—Decentralized

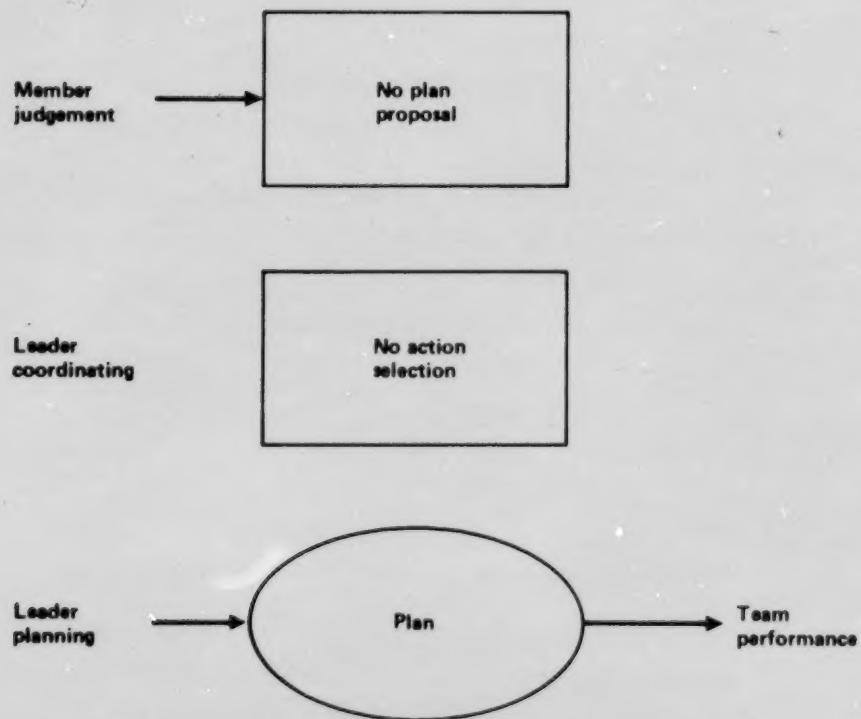


Fig. 7(b)—Centralized

This conclusion may make it seem as if a great deal of mumbo-jumbo has been offered to justify a common-sense conclusion. I think (hope) that this is too harsh. So long as we confine ourselves to considering problem solving and team performance *in the abstract*, as has been done here, the final conclusions are bound to sound vacuous. What I do propose, though, is that the methods described here could be instantiated. Given a specific problem, one could conduct a cognitive science analysis, perhaps even including computer simulation, to determine what the optimum team design would be. This would at best be a rough approximation. We could then set various teams to work on the problem and use the techniques of path analysis to evaluate the models and path coefficients required to describe their behavior. This would provide an orderly way of accumulating knowledge that would be of practical importance and that might point the way toward a uniting of concepts in cognitive science and organizational behavior.

DISCUSSION OF SESSION 4

Thorndyke commented that the analysis-of-covariance framework advocated by Hunt was problematic because the number of testable models increases exponentially with increasing numbers of independent variables. Hunt advised beginning with the simplest possible models and adding complexity only when these models cannot account for the data.

Burnstein asked whether Hunt's team model would apply when group members do not agree on the identity of the leader. Hunt replied that the model will "induce" an executive, indicating where the management functions of the team reside. Edwards objected that he knew of no team that merely gathers and integrates information. He asked whether tasks such as maintenance and engineering could be handled within Hunt's model. Hunt maintained that issuing commands, monitoring performance, and all other team functions except physical execution of commands could be incorporated into the model. He claimed that the model could be elaborated almost indefinitely by continually refining hypotheses and fits of the model to the data.

Menchik noted that a technique similar to analysis of covariance structures, called econometric analysis, was already in use in the study of some organizations. He advocated this procedure over analysis of covariance structures because it allows examination of reciprocal influences. For example, morale probably affects performance, but performance may also affect morale. Hunt claimed that analysis of covariance structures could be extended in this direction by using measures taken at several points in time.

Rick Hayes-Roth shifted the discussion to the goal analysis recommended by Collins. He asked if there were a method to identify and reduce the disparity between a team's nominal and actual goals. Collins advocated careful scrutiny of actual teams in operation to determine the nature of these disparities, followed by analysis and testing of structural variables that might contribute to these disparities.

VI. TRAINING AND INSTRUCTION

AN ANSWER TO TEAM TRAINING INSTRUCTIONAL TECHNIQUES: THE DESIGN OF NEEDS ASSESSMENT AND EVALUATION PROCEDURES

Irwin L. Goldstein

Since my expertise does not include the design of particular types of instructional techniques, I do not intend to focus in this paper on particular aspects of team training but rather on several topics which I feel underlie most of our problems in training (including what we will certainly eventually face in team training).

There are two apparently simple but actually very complex points that will be the focus of this paper. They are:

1. The development of the techniques and the evaluation model must be based upon a careful and thorough needs assessment of the organization.
2. The systematic development of any training or change techniques is dependent upon thoughtful evaluation which provides the information necessary to suggest revisions in our techniques.

ASSESSMENT OF INSTRUCTIONAL NEED

The needs assessment phase is the most critical aspect of the design of any change program. The choice of the program and the criteria chosen to evaluate the program all stem from the needs assessment. Psychologists are somewhat unique in their treatment of this step. For some reason, they have tended to focus on the technique rather than on a consideration of the needs and a determination of which technique fits which need. This has resulted in a fads philosophy of jumping from one technique to another. Yet we still don't know what technique works for which behaviors. Machinists don't choose their tools before examining the job, and a gardener usually chooses a sprinkling system rather than a bucket to water a half-acre lawn. Nevertheless, we still have to be warned by quotes like the following:

If you don't have a gadget called a teaching machine, don't get one. Don't buy one; don't borrow one; don't steal one. If you have such a gadget, get rid of it. Don't give it away, for someone else might use it. This is a most practical rule, based on empirical facts from considerable observation. If you begin with a device of any kind, you will try to develop the teaching program to fit the device (Gilbert, 1960, p. 478).

Gilbert is not saying that teaching machines or sensitivity training or any other techniques do not work. He is saying that the design of change programs does not begin with instructional media. Instead, we must, through needs assessment, determine the objectives of our programs so that our criteria for evaluation and the choice of instructional program are made realistically.

Having just completed the 1980 *Annual Review of Psychology* chapter on Training in Work Organizations (Goldstein, 1980), I certainly cannot indicate that team

training approaches are a growing fad. I went through over 4,000 references from an NTIS search, Psychological Abstracts, and ERIC search and reviewed hundreds of documents sent in by individuals. Yet the chapter has virtually no references to team training. Thus, the question becomes, why worry about fads? Actually, the answer to that is quite simple: Whenever we approach a new training issue, we always arrive at a solution which leads us to focus on the design of an instructional technique. The recent history of concern over interpersonal relations resulted in the T-group phenomenon. The focus on managerial behavior is centered around behavioral role modeling and the concern over pilot training has resulted in more and more dollars being spent on simulators. Yet T-group training has produced tremendous concern over its fit in the organizations; role modeling may be popular, but virtually no one knows what a manager does. Also, evaluation of Air Force simulators has not been fruitful. As noted by several researchers (e.g., Williges, Roscoe, and Williges, 1972), there is a lack of agreement on what constitutes terminal pilot performance and there is poor reliability for rating systems. (As an aside and a hint of my comments on evaluation, they also note that measures of retention beyond immediate performance in the instructional program are virtually nonexistent.)

My fear, of course, is that a big push on team training will lead down the same circular road with all sorts of emphasis on instructional techniques and no understanding of needs assessment, interventions in organizations, and evaluation models. If that is the case, it will go the way of all fads. John Flanagan (1973) said it well: "Recent experience indicates that the quality of the present educational programs can be improved more by systematic selection of what is to be taught than by improving how it is taught."

I will resist the temptation to describe the large number of training program failures which have occurred as a result of a lack of concentration on needs assessment and evaluation. In a sense, that would be like telling war stories. On the other hand, I am sure that sponsors would be a lot more careful about where they spend their money. Nevertheless, I will resist the temptation. Given all of this, what recommendations can be gleaned from the literature?

First, I will discuss needs assessment. Most needs assessment schemes refer to McGehee and Thayer's (1961) conception of organizational analysis, task analysis, and person analysis, which was updated in my own text (Goldstein, 1974).

Clearly, the specification of tasks to be learned provides critical input to the design of the instructional process. Interestingly, because of the emphasis on the instructional device, there is little attention paid to how to design task analysis systems as input to training. For example, there has been virtually no consideration of systems that examine the total job and provide information on which activities should be learned in training and which activities should be learned on the job. Similarly, no established procedures exist that empirically establish the content validity of a training program based upon a match of relevant tasks on the job and in the training program.

An important step forward is the work of Ammerman and his colleagues (1974, 1975, 1977). They provide a methodology for the development of tasks and for the design of scaling systems to measure a number of task dimensions. Among these are frequency of task occurrence, importance of tasks, and where and when the task should be learned. Their findings enabled them to decipher clues which indicated the need for training. Their data could also stimulate some important re-

search about the relationship between the various procedures used to scale tasks. In a study of auto mechanics they discovered a number of provocative facts, such as:

- Certain tasks were not performed on the job although supervisors said the tasks were part of the job.
- There were conflicts in the rating of tasks between supervisors and employees, between frequency and importance ratings, etc. These conflicts will have implications for later comments on organizational analysis.
- Tasks were being taught in training that everyone agreed did not belong as part of the entry-level skill and, vice versa, some entry-level skills were not taught in training.
- For many tasks, nontraining solutions (e.g., task redesign) were suggested.

Similar analyses of other jobs have also generated some interesting perspectives. Many observers (e.g., Campbell, 1978) have noted that if you want to know something about leadership you should observe what leaders do when they lead. A recent study of managers (McCall, Morrison, and Hannan, 1978) produced the following observations:

- Most activities were very brief.
- There were a large number of interruptions.
- Discontinuity is characteristic—i.e., tasks are handled in rapid-fire fashion, interspersing significant and trivial tasks without any special order.
- There were many types of activities: paperwork, phone calls, scheduled meetings, unscheduled informal meetings, and inspection tours and visits.
- There was a change in activities as the manager moved up the ranks.

Clearly, we do not have much information about these tasks, but my impression of the literature is that we know a lot more about what managers do than we know about teams. I am not talking about what teams do in the laboratory; that has limited relevance. I am talking about teams in work organizations—which should be what we are examining.

Similar comments could be made about our lack of knowledge about person analysis; instead I will mention one study and leave the rest to your imagination. Laabs, Panell, and Pickering (1977) developed performance-oriented tests to diagnose deficiencies in job performance for U.S. Navy missile technicians. That study found that performance deficiencies in maintenance skills were apparently related to the fact that the technicians did not get to practice their skills because of the extreme reliability of the missile test and readiness equipment. Also, the instructional programs were not particularly successful because the training was viewed as nonessential to the job, and thus many individuals never completed the instructional packages. This is simply another demonstration that needs assessments which describe what teams do when they are performing is a first priority.

On the basis of needs assessment techniques, it should be possible to determine what tasks are performed, what behaviors are essential to the performance of those tasks, what type of learning is necessary to acquire those behaviors, and what type of instructional content is best suited to accomplish that type of learning. Unfortu-

nately, these steps remain the most elusive aspect of the design of training systems, particularly for complex jobs.

There is another aspect to needs assessment, known as organizational analysis. I have already suggested that there should be concern about the fit between the training program and what is accepted in the organization. There are some of us who believe that perfect training might be accomplished and still there will be little transfer to the actual job.

Several authors have lamented that training programs are often judged to be failures because of organizational constraints that were not originally intended to be addressed by the instructional program. Salinger (1973) reported on case studies of six large federal agencies and discovered disincentives to effective employee training, such as supervisors who do not accept the practices taught by the training program or supervisors who cannot meet their required production norms while their employees are in training. She also found that:

- The benefits of training were not clear to top management.
- Managers do not account for training and development in production planning.
- Training programs external to the agencies taught methods contrary to the practices of the participant organization.

Another study (Baumgartel and Jeanpierre, 1972) noted that managers only utilized skills developed in training when the organizational climate was favorable. This was especially true for low-status managers. Allen and Silverzweig (1976) approach the disincentive issue by discussing the influence of group norms on training effectiveness. They suggested that it is necessary to determine the effects of behaviors such as bringing a new way of performing a task back to the work setting. Negative reactions from supervisors or other employees should make most training analysts reconsider their instructional plans. Research is needed to develop an understanding of the variables that could predict the likelihood of positive transfer from training to the on-the-job environment. We might ask the following questions:

1. Are there unspecified organizational goals which should be translated into training objectives or criteria?
2. Are the various levels in the organization committed to the training objectives?
3. Have the various levels and/or interacting units in the organization participated in the development of the program beginning at the needs assessment?
4. Are key personnel ready to both accept the behavior of the trainees and serve as models of the appropriate behavior?
5. Will trainees be rewarded on the job for the appropriate learned behavior?
6. Is training being utilized as a way of overcoming other organization problems or organizational conflicts that require other types of solutions?

If the responses to these questions leave an uneasy feeling of uncertainty, perhaps the training program should be reconsidered by performing an organizational analysis first. This implies that the best designed team training program may not result in transfer to the job. It is necessary to understand the constraints which

will prevent transfer of training. It also implies that to understand these issues one must examine them in work organizations. As an aside, this view also suggests that the separation of training as a technology from the study of organizational behavior is self-defeating.

EVALUATION

Evaluation is the systematic collection of descriptive and judgmental information necessary to make effective training decisions related to the selection, adoption, value, and modification of various instructional activities. The objectives of instructional programs reflect numerous goals ranging from trainee progress to organizational goals. From this perspective, evaluation is an information-gathering technique that cannot possibly result in decisions that categorize programs as good or bad. Rather, evaluation should capture the dynamic flavor of the training.

Efforts at evaluation appear to be evolving through a series of stages. In the most primitive stage, appropriate methodology is ignored and decisions are based upon anecdotal trainee and trainer reactions. While most analysts realize the limitation of this approach and apologize, there are still far too many efforts of this type. In the next stage, which could be just as unproductive in providing relevant information, the evaluation strategy is dependent upon strict adherence to the basic experimental methodologies of academic laboratories. This stage is characterized by designs which do not recognize the constraints imposed by the environment or the influence of the multitude of organizational variables. Often researchers find that these types of studies come to a screeching halt because of organizational constraints or because a technology has been applied which does not answer the questions being asked. A good illustration is the use of traditional "do nothing" control groups in situations where that is clearly not the appropriate comparison. Since the researcher has faith in the results precisely because experimental methodology has been employed, these types of efforts can be particularly misleading. A third stage is observed when methodologies are matched to the constraints of the environment. It is characterized by careful consideration of threats to validity and the creative application of design methodology to the questions being investigated. A fourth stage which interacts closely with the third stage is the development and understanding of the philosophy of intervention. The work here recognizes that the instructional program itself, including the evaluation effort, interacts with the organization to produce a variety of effects. An illustration of these issues is provided by Hand and Slocum's (1972) study, which speculated that the decrements in control group performance occurred as a consequence of their being upset over the results of the more favorable treatment given to the training group.

From a design point of view, it is necessary to be much more creative in the development of evaluation models. Those models should permit the extraction of the greatest amount of information within the constraints of the environment. Thus, it is important to continue to develop information about constraints that threaten the understanding of data collected in organization environments. However, it is just as necessary to design models that allow the collection of data with maximum confidence. Researchers cannot afford to be frozen into inactivity by the spectre of threats to validity. It would be helpful to have further information about alternative evaluation models, including the use of individual difference method-

ology and content validity strategies. Hopefully, these models will emphasize the relationship between training and on-the-job performance as well as the examination of learning performance in the training program. This latter concern raises the question of the criterion issue.

Appropriately, the main concern in criterion development has been procedures for the establishment of relevant criteria that reflect the dynamic nature of the training process. Many researchers (e.g., Freeberg, 1976) have noted that the critical dimensions for criterion development are the need for multiple behavioral indices, the specification of the criteria along the temporal continuum represented by the training process, and the fundamental requirement of relevance. The difficulties in establishing appropriate criteria for individual performance argue that energetic all-out efforts are needed if we wish to understand team performance. For example, the application of criterion-referenced testing procedures to training settings is a procedure based upon concern for establishing relevant criteria. As discussed by Glaser and Nitko (1971), criterion-referenced tests are planned to be representative of a specific domain of tasks. Thus, criterion-referenced measures are designed to compare an individual's test score with a standard, while norm-referenced tests are designed to compare a person's score with the performance of others. Several thoughtful discussions concerning the development and interpretation of criterion-referenced measures have been published (Swezey, Pearlstein, and Ton, 1974). Unfortunately, an inspection of the literature on the use of criterion-referenced measures in training situations indicates that the term is more a philosophy of hope than a reality in the development of relevant criteria. Normative strategies have been criticized because they do not insure that the trainees can perform tasks at a mastery level. However, it is sadly apparent that the problems of translating information from the needs assessment into instructional objectives and relevant criteria remain whether or not the measures are called criterion-referenced. While this reviewer strongly supports the goals of criterion-referenced measures, most of the training work is governed by excess verbiage with little empirical support. In a rare exception, Pannell and Laabs (1979) examined a methodology to evaluate a criterion-referenced test. First, they selected test items that both showed a significantly greater proportion of correct responses for a post-instruction group and were answered correctly by at least 50 percent of the post-instruction group. Then they used the items on a hold-out sample to see if this would enable them to predict whether the trainees were members of a pre-instruction group. After refinement of the items, the investigators predicted pre- and post-instructional members with 68 percent to 92 percent accuracy for the different modules. More research focused on this type of approach is badly needed.

An examination of many of the above studies supports J. P. Campbell's (1978) concern that the field is in danger of being swamped by questionnaire-type items. The failure to develop methodologies for systematic observation of behavior is a serious failing. The importance of establishing multiple criteria that reflect the observing of key behavior seems a lost art. One important development may be D. T. Campbell's (1974) process-variable approach, which is used to collect qualitative information to help further understand the processes of intervention as well as the meaning of the quantitative indices.

Campbell notes that our evaluation efforts have emphasized outcome and ignored process measures. Thus, evaluation efforts have missed the richness of detail concerning how events occurred and even what went wrong. There is other re-

search on studies (e.g., Goldstein, 1978) where data misinterpretations would have occurred if the investigators had been unaware of the events that occurred during training.

In summary, I would recommend concentrating on the following:

1. Develop needs assessment techniques that provide input about what teams do when they perform as teams.
2. Emphasize research that examines the fit between training and the organization where it is going to be transferred.
3. Develop relevant multiple criteria that reflect team performance.
4. Develop creative evaluation designs that can be used to examine performance in real work organizations.

I do not want to close without saying that someone should develop a team training instructional approach. That is not a problem, though, since we know that the fads approach will get around to that. Let us at least demand high-quality empirical investigations that examine the usefulness of all of these techniques which seem so promising. Let us make sure that someone explores the many boundary conditions for training studies. For example, most investigations of flight training, rater training, and business games employ undergraduate students as subjects. Even when studies are conducted in work organizations, they are usually employed in a research-only design. It might be informative to explore the differences between these conditions and the use of procedures in an actual organizational system.

FROM PERSONAL TO COOPERATIVE COMPUTING

Ira Goldstein

INTRODUCTION

In this paper, I contrast the present paradigm of personal computing with an alternative paradigm that emphasizes cooperative design. I believe that understanding this second paradigm is central to creating computing environments that support team problem solving.

Personal computing is the wave of the future, at least as announced in the many magazines that have sprouted on this subject. Indeed, the mass availability of computing cycles will certainly have a significant effect on our society. However, the software environments being offered under the banner of personal computing are usually only cut-rate versions of the standard environments of the last decade. The revolution in personal computing is at present a manifestation of a hardware revolution, not a software revolution. From a software standpoint, the personal computing bandwagon is obscuring what I believe to be a central research problem of the 1980s—the development of cooperative computing environments.

Developing computing environments that are congenial to cooperative computing is important because present software systems, which are generally developed by teams of programmers, are becoming increasingly complex, yet our tools for coping with this complexity are lagging far behind. Evidence for this claim is based on the widespread complaints regarding the software problem, i.e., that the cost of developing and maintaining today's software is out of control.

TWO PROGRAMMING PARADIGMS

The personal computing paradigm in admittedly oversimplified form consists of the following sequence of steps:

Define → Explore → Debug → File

The programmer begins by defining his program, then exploring its behavior. This is followed by debugging, if necessary, and concludes with the filing of the debugged code.

Let us contrast this cycle of individual design with an alternative sequence emphasizing cooperative design:

Retrieve → Adapt → Explore → Debug → Describe → Store

In cooperative problem solving, a designer begins more frequently with the retrieval of related programs than with the construction of an entirely new program. The designer is a member of a large community and it is likely that someone in that community has solved a related problem. The most efficient beginning is therefore to adapt that solution to the present problem. It may be that no one has considered a relevant problem. However, it is clearly absurd to ignore previous experience.

Exploration and debugging of proposed solutions remain common to both paradigms. However, the final stages of the problem solving are different. To support subsequent use by others, solutions in the cooperative computing paradigm are not complete until properly documented. To be understandable by others, this documentation should contain a discussion of the underlying design decisions. Such documentation is generally missing in personal computing, since it is rare that anyone besides the creator of the code examines the listings.

The last step in the personal computing paradigm is to file the code. Subsequent retrieval is accomplished by the use of a file name of a few characters. This is an adequate mnemonic key for an individual problem solver, but it is clearly insufficient when used as an entry description by others. Like a library, storage in the cooperative computing paradigm involves indexing the results under a wide variety of keys.

AN EXAMPLE

To clarify the differences between these two paradigms, I have selected a typical student exercise drawn from the Logo computing environment. I have chosen this environment since the Logo project was among the leaders in the early seventies in developing educational computing.

The goal of the exercise is to define a program that causes a graphics cursor called the Turtle to draw a triangle. Below is a typical program that a student might initially produce:

TO TRIANGLE	
10 FORWARD 100	"Move 100 in the direction you are pointing."
20 RIGHT 60	"Turn 60 degrees to the right."
30 TRIANGLE	"Recurse"
END.	

The student would then be encouraged to explore the behavior of his program. In this case, the program draws a hexagon, since the rotation corresponds to the external, not the internal angle, of the the figure being drawn. Subsequent debugging would result in the student changing the rotation of line 20 from 60 degrees to 120 degrees. With this correction, the program is successful in the sense that a triangle appears on the screen, although the program does not terminate. The student might or might not choose to debug the infinite recursion. The session is completed by filing the program.

This scenario is successful in exercising skills of planning and debugging needed by an individual when engaged in personal problem solving. However, it is not successful in exercising the additional skills needed when that individual is a member of a team. This scenario demands no skill in asking the appropriate questions of the community data base or in describing a proposed solution in such a fashion that it can be usefully adapted by others in the future.

Were this scenario to take place in an environment that emphasizes cooperative computing, the program drawing the triangle would be only the kernel of a larger set of descriptions produced by the programmer. This set of descriptions would contain the problem, the constraints, the caveats, and the plan. For example, the solution might take the following form:

PROGRAM: <as above>
 PROBLEM: "Produce a program that draws a triangle."
 PLAN: "Simplify. Produce a program that draws equilateral triangles."
 CAVEAT: "The rotation is the external, not the internal angle."
 COMMENT: "The program does not halt."

It is this bundle of descriptions that is stored in the community data base. The programmer should provide as many keys as possible, possibly indicating that it should be cataloged under equilateral triangle programs, with additional keys placed under polygon and triangle programs. The community member faced with a related problem and inquiring under any of these keys could examine the program along with its rationale. This set of descriptions would then become his or her starting point.

TECHNICAL ISSUES

To develop cooperative computing environments, a host of technical issues must be addressed. In this section, I outline several.

1. Research on data bases and question-answering must be applied to the storage and retrieval problems. Listings, the hallowed method of old for documenting programs, will no longer be adequate. Programs must be treated as objects in a data base, retrieved by various queries and stored under multiple indices.

2. Research on knowledge representation languages must be applied to the description of programs. Such languages are necessary to express the intent, plan, and properties of a program in a fashion that can be subsequently queried. However, the scope of a cooperative computing environment is too large for any current knowledge representation language to allow all descriptions to be formalized. Therefore, contrary to the usual research on knowledge representation, it will be necessary to develop a formal descriptive system that can gracefully degrade to informal text, when the formalism proves inadequate.

3. Research on man-machine communication strategies must be applied to prevent the programmer from drowning in a sea of irrelevant data. As the community data base grows, the need to filter the information presented to a programmer becomes more crucial. There is no single means to achieve this, but there are many strategies that can be applied. These include user modeling, explanation strategies, high-bandwidth displays, and multimodality output.

4. Research on parallel computing must be applied to cooperative environments where more than one member of the community may be examining and modifying some part of a joint design. The traditional computer science problems of resource conflict and deadlock have been studied for multiple processes. These studies must be extended to encompass the requirements of an electronically linked community of human problem solvers.

5. Research on achieving an integrated environment for each individual must be applied, since a cooperative computing environment inextricably links programming to text editing, message sending, and information retrieval. The text editing

is required to supply the necessary documentation. Message sending is the channel of communication between community members. Information storage and retrieval form the basis of the transactions between the individual and the community library.

6. Research on computer assisted instruction must be applied to making the complexities of cooperative computing environments understandable to new users. The complexities of community storage and retrieval are the price of successful cooperation. But the price is too high if a new team member cannot learn how to employ these mechanisms successfully. A machine consultant within the cooperative environment can supply documentation about the environment itself, maintain a user model to filter this documentation, and employ explanation strategies and a task model to deliver its documentation at appropriate times.

FIELD STUDIES

As cooperative computing environments are developed, field studies will be crucial if we are to understand the sociology of these environments. Among the questions that will need to be examined are: When problem solving acquires this cooperative flavor, how is credit to be apportioned? How can the community cope with members whose contributions have been found to be flawed in the past? Will there be problems of social isolation when the community members are non-local, linked only by electronic means?

RESEARCH PROGRAM

The most critical need at this time is to create a number of cooperative computing environments. Neither the technical nor the sociological issues can be studied in the absence of such environments. The following are criteria for the placement of such an environment:

1. The communities chosen for these environments should be real ones in the sense that they have a need for the environment and employ it on a daily basis.
2. The members of the community should be open to a new approach to cooperative problem solving.
3. Each member of these communities should be provided with the level of computing cycles that will be common at the end of the 1980s.
4. This research is sufficiently new that these environments should be sheltered.

These considerations point to placing the first cooperative computing environments into the organizations doing the research. These organizations engage in team programming but are neither very large nor exposed to all of the tumult of the operational world. By including the researchers, an important feedback loop will be created to encourage their development. By including other parts of the researchers' organization, a counterbalancing check on the environments can be obtained.

DISCUSSION OF SESSION 5

Rick Hayes-Roth began by noting that technological innovations such as text-editing and message-handling systems can ease the strain caused by manpower shortages in the military. Martins warned that individuals at high levels in the military often oppose automation because they fear that centralization of control may jeopardize their command prerogatives. However, it was generally agreed that a technological revolution is under way in the military and that in the near future technology will drastically change current procedures. Ira Goldstein admitted his concern at the use of computers for strategic and tactical planning and decisionmaking. However, he emphasized the immediate need for computer applications in the areas of communications, text handling, and information retrieval.

Weiner expressed concern that many team tasks are not sufficiently stable or comprehensible to support task analysis. He also suggested that an overly specific task description would limit the flexibility of a team and reduce its ability to respond to clever countermeasures. However, Ira Goldstein replied that task analyses and training curricula can support flexibility if that is a desired goal. Roberts noted that organizational research has begun to recognize the need for flexibility, for example, in the concept of the "experimenting manager."

Hunt raised several potential problems with automated message systems. He noted that it is difficult to convey subtleties such as indirect speech acts using a computer message system. He also argued that the loss of the non-verbal linguistic cues (e.g., authority conveyed by expression, intonation) available in face-to-face communication would greatly increase the possibilities of misinterpretation. Weiner replied that individual styles rapidly emerge in computer message systems. However, Edwards pointed out that today we still use face-to-face communications for important meetings, even when we could use the telephone. Ira Goldstein noted that communications systems of the future could include audio channels. Research must investigate the utility and requirements for audio communication.

VII. HEURISTIC MODELING

TEAMS AND DISTRIBUTED ARTIFICIAL INTELLIGENCE

Nils J. Nilsson

INTRODUCTION

Technical developments in artificial intelligence (AI) have supported information processing theories of human cognition which are permeating modern psychology. Thus, AI has produced vision systems, speech understanding systems, medical diagnosis systems, robot systems, and other computer-based mechanisms that display various cognitive skills. These systems, and the theories on which they are based, have provided metaphors, vocabulary, inspiration, existence proofs, and concepts useful for psychologists in constructing mechanistic models of cognition.

Although AI has made impressive progress in building *individual* cognitive systems, it is only recently beginning to consider the problem of building *teams* of systems. In this note, we briefly mention some of the AI research directed at *distributed artificial intelligence*, discuss some of the key research problems, and describe some initial work in this area now being started at SRI. Just as previous AI work on individual systems helped fuel theoretical developments in the psychology of cognition in the individual, it can be expected that work in distributed AI will inspire new and important theoretical work in the psychology of teams.

SUMMARY OF AI RESEARCH RELEVANT TO TEAMS

Work in Distributed and/or Parallel Processing

Several developments in AI are under way that may be relevant to building a theory of teams. One is the "actor" formalism of Hewitt (1977). Hewitt proposes that computation be performed by a large number of units that send "messages" to one another. To date, Hewitt has concentrated on relatively "low-level" computations, and it is unclear how useful his ideas would be as a model of the high-level interchanges that take place between members of the sorts of teams we have in mind.

Another important effort involves work of Lesser, Erman and their associates. Lesser and Erman (1979) describe a distributed HEARSAY system. (HEARSAY is a special AI system architecture that has been applied to speech understanding.) Lesser and Erman experiment with a version of HEARSAY that is cut up into components that could be located in several different physical locations. These communicating components then cooperate to solve problems such as speech understanding. Similarly, Corkill (1979) describes a distributed version of NOAH. (NOAH is a hierarchical planning system for generating robot plans.) Lesser and Corkill (1979) discuss other related projects in distributed AI.

Smith (1979) describes a "contract net" system in which tasks are parceled out among parallel operating subsystems. These subsystems communicate and cooperate in solving problems. Work is continuing at Carnegie-Mellon University, under the direction of Reddy, on parallel and distributed computation.

Research on Speech Acts

Recent research by Cohen (1978) and by Allen (1979) on speech acts appears to be very important in constructing a theory of teams. Following earlier ideas of Searle (1970), Cohen and Allen have developed systems for planning sequences of "speech acts" and for recognizing the plans that lead to speech acts. This work assumes that natural language communication is performed as part of a process of executing plans that accomplish the goals of the participants in such communication. Thus, if my goal is to get a door open, I might either perform the action of opening the door myself or I might perform the "action" of saying, "Please open the door." Such a statement is a "request"-type speech act. There are also "inform"-type acts and others. Planning speech acts requires that the planning system have a model of their preconditions and effects, and this model must be based on a method for reasoning about the knowledge and beliefs of others.

Appelt, a Stanford graduate student, is currently doing work at SRI on generating English sentences that accomplish given goals. Since communication is at the heart of the team process, recent AI work related to speech acts would appear to be highly relevant.

Reasoning About Knowledge and Belief

In order to function effectively, a team member must know what sorts of knowledge and skills are possessed by other team members. It (or he or she) must know about the roles (e.g., leader, specialists) of other team members and about its own role. Furthermore, it must know how to reason with this knowledge in order to make and execute effective plans and in order to generate and understand communication with the team. AI research is just beginning to make progress on the problem of representing and reasoning about the knowledge and beliefs of others. Among those who have suggested approaches to this problem are McCarthy (1977, 1979), Moore (1977, 1979), and Weyhrauch (forthcoming). The interaction between this work and the speech-act work is of the utmost importance in building teams of AI systems.

SRI DISTRIBUTED ROBOTS PROJECT

We are just beginning work at SRI on a research project that will involve a team of (3, 4, or 5) robots. The project will be supported by the Information Systems Program of ONR. Besides myself, we expect that the following people will participate in the project: Kurt Konolige, Bob Moore, Earl Sacerdoti, and Dave Wilkins.

Imagine a team of robots that cooperates to solve a task such as building a tower of blocks. (We have not yet decided finally on the task domain for our robot team, but many important problems arise even in the well-visited "blocks world.") We could view the team of robots as merely a set of several effectors all under the control of one AI planning/perception/execution system—such as the SRI Shakey system of a few years ago—or we could view the team as several more-or-less independent robot systems, each with its own planning/perception/execution systems.

We note that teams of independent robots require less interrobot channel capacity, are less vulnerable to failure of individual robots, and can more easily be

augmented by additional robots. Such factors may well be of critical importance in many applications. In this project, we assume the latter design and accept the consequent problems of how the individual robots are to model each other and communicate with each other.

Our work is beginning by considering the following research problems:

1. How should each robot represent its own goals, its plans for achieving those goals, and its progress in the execution of such plans? (Previous representational schemes, such as those used in STRIPS, say, were adequate for representing actions and configurations of physical objects. These schemes did not really address the problem of representing goals and actions in a manner in which they could be reasoned about.)
2. How should each robot represent the existence, beliefs or knowledge, and capabilities of the other robots (including, in turn, similar knowledge that these other robots might have, and so on)?
3. How should each robot make and execute plans that take full account of the helpful abilities of other robots?
4. How should the robots communicate with each other? (Note that if robot S1 wants robot S2 to know "B," and if S1 knows that S2 knows both "A" and "A implies B," and S1 knows that S2 always infers anything inferable by a one-step modus ponens operation, then S1 can cause S2 to know "B" by telling it "A." Communication efficiency is enhanced by using strategies such as this because "A" may be a much shorter message than "B." Natural language communication involves many of these same kinds of inferences.)
5. What problem-solving and action-implementing architectures should be used for each robot system so that the effect of robot S1 telling robot S2 to achieve a goal actually results in S2 planning and executing actions to achieve that goal (if the goal is consistent with other goals, prohibitions, and knowledge that S2 has)? We are interested here in questions about the "dynamics" of the whole assemblage. Each component robot has to be organized in such a way that the sum of their individual perception, problem-solving, action, and communication processes results in appropriate performance of the team.

Our initial approach is to investigate Weyhrauch's (forthcoming) FOL system as a candidate architecture for each system. We hope that some simplified version of FOL might suffice. Whereas Weyhrauch has explored the use of a single FOL in various tasks, especially those that can profit from what he calls "metalevel" reasoning, we would want to have several FOL-like systems all operating simultaneously.

Perhaps each robot would consist of a rule-based deduction system with a data base of rules and predicate calculus fact expressions. The fact expressions might make statements about the physical world, such as, "Block A is on block B," or they may make statements about the data bases of other robots such as, "Robot S2 has in its data base the statement 'block A is on block B'." This latter expression is a way of saying, "Robot S2 believes that block A is on block B." Using additional nesting, there may be statements like, "Robot S2 has in its data base the statement 'robot S3 has in its data base the statement 'block A is on B'.'"

A key feature of FOL is its scheme for "procedural attachment." We could use procedural attachment to help perform reasoning tasks. A linguistic item in *S1*'s database, such as *S2*, could refer to the database of robot *S2*. Furthermore, robot *S1* can have a *partial model*, say *S2**, of the expressions in *S2*'s data base. *S2** could contain expressions that refer to acts that *S1* knows *S2* knows. Now when *S1* wants to reason about what *S2* knows, instead of manipulating facts of its own, it can, by procedural attachment, use a deduction system (modeling *S2*'s reasoning abilities) operating on *S2**. Moore (1979) points out several difficulties with this so-called database approach to reasoning about knowledge. The procedural attachment scheme lets us use the database approach when it escapes these difficulties, but it allows us to fall back on stronger methods (involving metalevel reasoning about the knowledge of others) when stronger methods are needed.

Even though this work is just beginning, we expect that it will lead to advances in methods for reasoning about knowledge and belief and in methods for understanding and generating natural language statements. It should also enhance our theoretical understanding of group problem-solving processes and of the dynamics of both formal and informal groups.

THE FUTURE OF HEURISTIC MODELING IN TEAM TRAINING APPLICATIONS

Gary R. Martins

HEURISTIC MODELING: WHAT IS IT, HOW IS IT DONE, WHAT IS IT GOOD FOR?

Recent advances in computing have significantly expanded the set of tools and methods available to model builders. Classical computing techniques have long been available for *procedural* modeling, using numerical approaches to represent the state and behavior of the modeled entities. Recent work in the field of artificial intelligence is preparing the way for a new class of models, appropriate to new kinds of users and applications. These *heuristic* modeling techniques employ non-procedural syntactic forms and natural language expressions for representing knowledge. The new class of models is likely to have a significant impact on the design, training, and evaluation of teams, and other organizations. This tutorial paper first defines and gives examples of heuristic modeling, then describes the process of building a heuristic model. A number of applications of obvious relevance to the study of teams are then described. We close with some recommendations for future research and development.

A Definition and Some Successful Examples

Heuristic modeling may be defined briefly as the building of human-like expertise into computing systems. Human experts often rely on knowledge that is fuzzy or incomplete and reasoning that gives only approximate solutions; however, this combination of partial knowledge with approximate reasoning can be very effective for solving ill-structured problems in complex domains. Heuristic modeling attempts to embody this sort of powerful knowledge in a computer program. From its beginnings, the field of *artificial intelligence* has been concerned with modeling human expertise of various kinds, and that has pioneered the development of useful tools and methods. We can illustrate some successes of heuristic modeling with a few examples.

The HEARSAY-II system, developed at Carnegie-Mellon University, has demonstrated the feasibility of computer recognition of connected human speech. The original HEARSAY-II system nicely exemplifies the heuristic approach: the representational mechanisms provide great flexibility, and the internal structure of the system corresponds to a number of hypotheses about human cognitive functioning.

The PARRY system, developed at Stanford, also models human language performance, but with a different twist. PARRY engages in a textual conversation with its user, in the course of which it exhibits paranoid language symptoms; the degree of paranoia, from normal to extreme, is controlled by an external parameter setting.

The MYCIN and INTERNIST systems, developed at Stanford and the University of Pittsburgh, respectively, model the behavior of expert medical practitioners. The MYCIN system embodies expertise in the prescription of therapies for the

treatment of infectious blood diseases. INTERNIST is a physician's bedside assistant, guiding the doctor through an optimized sequence of diagnostic tests and advising him about the interpretation of results. When their recommendations seem surprising or doubtful to the user, both the MYCIN and INTERNIST systems are able to explain the chain of evidence and inference supporting the recommendations.

The TECA system embodies the naval officer's expertise in assessing the evolution of a tactical ocean surface warfare threat. TECA was developed and maintained by actual Navy users at the Naval Ocean Systems Center, rather than by computer experts. TECA thus illustrates an important goal of artificial intelligence: to make it possible for end-users to create sophisticated software of their own.

The Art of Heuristic Modeling

There are several ways in which heuristic models can be built. Figure 1 depicts three approaches. The path at the left represents the classical approach to model building. Here, a specialist trained in the art of model building observes the expert at work. Analysis of verbal protocols supplied by the expert supplements the specialist's own observations. From these sources, an analysis of subtasks is derived and progressively refined to produce a model—more accurately, a description of a model. This distinction highlights a key difference between a classical descriptive model of some expert behavior and a computerized heuristic model of the same behavior: The former *describes* the behavior, while the latter actually *reproduces* it.

Another approach is shown in the center of the figure. Here an expert produces a system that mimics his own competence. This expert modeler needs a good modeling language, whose level and form of expression are well matched to what he knows and the manner in which he naturally tends to express that knowledge. Hence, one major concern in the field of artificial intelligence has been to build modeling languages and other tools that help meet this need.

The right-hand path in the figure depicts a methodology for model building that requires a minimum of human intervention. In this emerging paradigm, expert human behavior is observed and analyzed by a sophisticated computer program, which attempts to duplicate and then generalize from the observed behavior.

The Science of Heuristic Modeling

Once the model builder's *artistry* has yielded a system whose performance is in some sense interesting, the model may yet be subject to substantial improvements by re-engineering its implementation. This is the *science* of model building; its goal is to adapt the model to its new electronic environment, producing a system that is effective, efficient, and affordable, without changing its essential behavior. Moreover, a properly engineered model:

1. Has a good user interface.
2. Has the versatility to cover the full range of problems to which it has been addressed.
3. Presents no daunting problems of maintenance, upgrading in the field, and integration with companion systems.

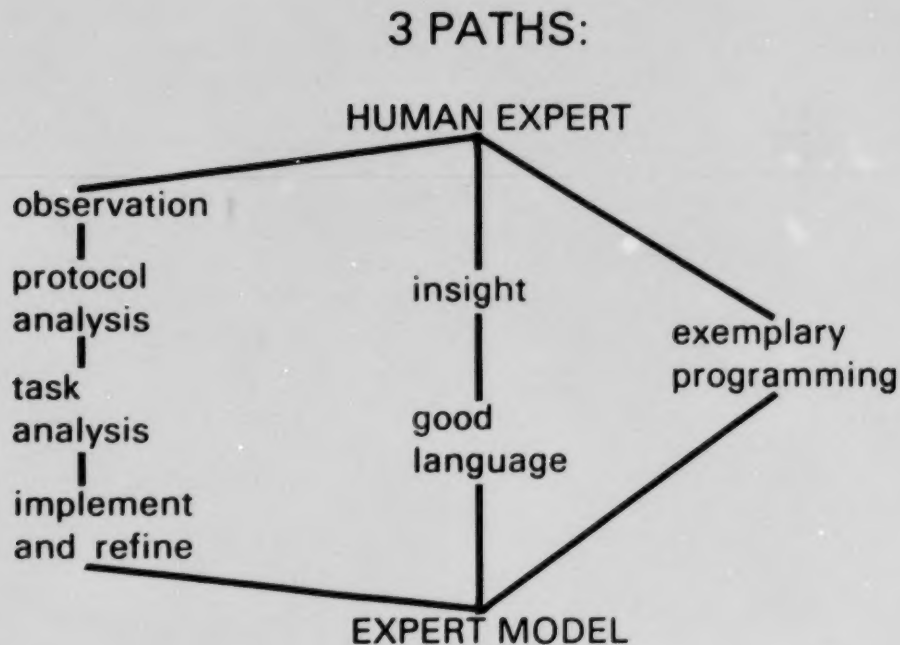


Fig. 1—The art of heuristic model building

A major reason for optimism in viewing the future of heuristic modeling is the substantial improvement in the quality of the tools available to model builders. FORTRAN was, of course, the first high-level language used for modeling. The early models were heavily numerical and *procedural* and were used primarily for the representation of physical systems and in operations research problems. The SIMSCRIPT and SIMULA languages, procedural languages specifically designed for simulation, incorporated numerous specific features not present in FORTRAN to help programmers with modeling applications.

The advent of LISP and its derivatives (PLANNER, DIRECTOR, etc.) greatly expanded the range of control structures and representational forms available for the programmer. For example, they facilitated the representation of objects, attributes, and values. These *non-procedural* languages were used to build a new class of sophisticated models, often aimed at duplicating various aspects of human expertise.

A new class of software modeling tools is aimed not at programmers but at experts in other fields. Using these new software tools, non-programmer experts can create their own machine assistants or counterparts for many purposes. The RITA system was developed at Rand in 1976. Based upon the production-rule representational scheme of MYCIN, RITA was not limited to a single field of application. Using RITA, experts at the Naval Ocean Systems Center built TECA—their own model of naval surface tactical threat evolution. Other RITA users have built models for network protocol testing, database query management, military command post simulation, and other applications. Rand is now building ROSIE

(Rule-Oriented System for Implementing Expertise), a second-generation production-rule language that offers significantly greater power and flexibility to the model builder, while remaining usable by non-programmers.

ROSIE: A Tool for Building Heuristic Models.

A ROSIE model consists of two different kinds of entities. First of all, there is a set of *objects*, each of which has a number of explicit attributes; the values of the attributes may be numbers, textual strings, or pointers to other objects or sets of objects. ROSIE objects may represent concrete, physical entities, such as people or locations or equipment. They may also represent more abstract entities, such as plans, organizational structures, goals, and so on.

The behavior of objects is represented in a collection of *production rules*, the second major component of a ROSIE model. A production rule has two parts: The first part describes a set of tests to be carried out on the model's database of objects; the second part describes a set of actions to be taken if the foregoing tests are satisfied. Production-rule syntax takes the form:

IF (tests are true), THEN (perform actions);

For example,

IF a quorum of members is present,
THEN the chairman begins the meeting;

Or,

IF all tasks of an assignment are completed,
THEN start the next assignment;

Or,

IF an attack is anticipated,
AND radar indicates an approaching plane,
AND the allegiance of the plane is not known,
THEN add the plane to the list of bogeys,
AND scramble interceptors to the plane,
AND set yellow alert;

And so on. English-like expressions are widely used in ROSIE models to specify the attributes and values that attach to objects and to describe both the conditions and actions of production rules.

The ROSIE knowledge representation offers several key advantages over more traditional languages. The formal structure of production rules is simple and uniform. This, together with the natural language expression of conditions and actions, makes them more readily understandable to end-users. The contrast with equivalent FORTRAN code, for example, is striking!

Production rules provide a natural chunking of knowledge, in that each rule represents a relatively autonomous specification of behavior. The system's simplified syntax allows the user to manipulate the various production rules independently of one another. The model builder may experiment quite freely—adding, deleting, or changing rules—in pursuit of the desired overall behavior. Thus, a

heuristic model may be built up incrementally by experiment, rather than having to be designed in the usual monolithic fashion.

Finally, the rules which generate the models' performance themselves can provide a basis for the system to explain its own behavior to the user. This simplifies the subtle task of debugging and refining a heuristic model.

APPLICATIONS OF HEURISTIC MODELS TO TEAM PROBLEMS

We have just briefly reviewed the character of heuristic models, the ways in which such models get built, and the kinds of tools that are becoming available to model builders. Let us turn our attention now to the kinds of applications that heuristic models will find in the design and training of organizations and teams. In the following paragraphs, I should like to describe four classes of such applications:

1. The creation of an institutional memory.
2. Management of the training/gaming context.
3. Training management, including coaching of teams.
4. Organizational engineering.

Institutional Memory

Many organizations expend a large share of their resources on the acquisition and development of expertise of various kinds. A key problem for most organizations is the fragility of this corpus of expertise. An organization's pool of expertise resides in the heads of the people who are, for the time being, members of the organization. In general, this is the *only* place where this expert knowledge is stored. Thus, ordinary as well as catastrophic turnover of personnel poses a constant threat to this vital resource. Organizations have traditionally relied upon books of procedure, training manuals, courses of instruction, apprenticeship programs, and the like to combat this problem. While such media may sometimes embody good descriptions of expert behavior, they cannot directly exercise such behavior for productive or exemplary purposes. The possibility of effective and faithful computer models of expert performance offers a new and more effective kind of remedy. As such models are built—moving expert knowledge from people into machines—organizations can expect substantial benefits.

First of all, model-based expertise is permanent; it is always available, since it is not subject to retirement, vacations, or other absences. Its performance is reliable and predictable, and it is unaffected by work schedules, fatigue, emotional stress, or other adverse factors. Moreover, this expertise can be deployed in environments hostile to human performance.

Second, good computer-based models will share an important property of human expertise: incremental enhancement over time. The best human experts continue to improve their performance over the span of a career. Computer-based heuristic models can do likewise, refining and broadening their behavior both through internal learning mechanisms and through tuning by users. Since the "career" of a computer-based model has no intrinsic time limit, its performance could continue to improve indefinitely.

Third, heuristic models are explicit and overt. That is, their internal composition, down to the finest details, is always available for inspection and review. Such models are far more easily shared and compared than their human counterparts. The pedagogical advantages of this difference are significant and obvious.

Finally, computer-based models are economical. Such models may be costly to create in the first place, but the cost of training human experts—such as air traffic controllers or physicians, for example—is also considerable. Once such models are created, the cost of replicating, distributing, and maintaining them may be negligible. This substantial economy could be used to justify relatively large expenditures on the initial creation of heuristic models.

Training and Gaming Environments

A second major application of heuristic models for training purposes is the management of the training or gaming context. For many familiar reasons, it is necessary to embed the training experience in a realistic simulation of the intended operational environment. Heuristic models offer substantial advantages over the more traditional approaches to this requirement, in terms of cost, safety, and flexibility. Person-intensive approaches to environmental simulation involve problems of scheduling, cost, rehearsal, and control, among others. A command post exercise, for example, presently requires the services of control team personnel to generate and schedule inputs to the trainees, and to respond appropriately to their actions. A heuristic model of this same environment could assume most, if not all, of these functions, with some unique advantages for training effectiveness. These advantages include:

1. More consistent and repeatable environmental behavior.
2. Tighter scheduling and control than can generally be realized through control team discipline.
3. Ready availability.

Thus, for a comparable level of resource expenditure, training centers should be able to develop and deploy a wider variety of training contexts, with better performance management.

Heuristic models may also be applied in replicating the behavior of other players, both collaborative and hostile. A widespread problem in military command training is the difficulty of adequately representing the expected behavior of the enemy commander, both tactical and strategic. Career path problems may make this an unpopular assignment. Further, personnel who expect such an assignment rarely have access to the time or other resources necessary to master the role in depth. Clearly, a strong heuristic model of enemy behavior would alleviate the chronic shortage of trained personnel and would offer the prospect of much improved performance.

Training Management: Coaching, Scenario Control, and Evaluation

Timely and accurate feedback on performance is recognized as a powerful training device. Unfortunately, this approach has been difficult to apply consistent-

ly, except where individualized instruction is feasible (a luxurious situation that rarely occurs). A heuristic model of the desired behavior can provide an effective, vigilant, and inexpensive substitute for the expert instructor. The responses of the model to training stimuli can be compared with those of the student, providing a basis for reinforcement or correction. Most computer-aided instruction efforts are based upon this premise; the limited success of these efforts is due, in general, to the weakness of the behavior models they have employed. Heuristic models offer a unique advantage in the coaching application. The explanation capability of modern modeling systems, such as ROSIE, provides an immediate basis for responding to student questions and deepens the student's appreciation of his expected role.

Up to this point, we have stressed the model-building application of systems like ROSIE. It should not be forgotten, however, that ROSIE—and systems like it—can well serve many of the common functions of more conventional programming languages. Because of their strong end-user orientation, such systems may be preferred to more conventional programming languages for the control of scenario development and for the administration of complex tests.

Sophisticated performance assessment may be abetted by the ability to handle and evaluate complex testing procedures. Computers are often used to control the selection and sequencing of test materials. In addition, computerized test and scenario administration simplifies the collection of response data, the analysis of collected data, and the provision of immediate feedback to the student, where appropriate. Heuristic-programming systems should make it possible for training personnel to develop and administer sophisticated training scenarios and tests without programmer assistance. The benefits of this approach include more rapid materials development and tuning, greater freedom for experimentation, and reduced costs.

Organizational Engineering

Let us briefly consider one final application of heuristic models in the field of team development: *organizational engineering*. If we can build effective heuristic models of individual expert performance, then we should be able to model the performance of an organization by linking together individual models of its components. The resulting organizational model then may serve as a vehicle for experimentation and refinement in group structure.

In a network of expert models representing an organization, the component submodels may be characterized by answers to such questions as:

1. Who has what resources?
2. Who creates, sends, and receives information?
3. Who has what responsibilities?
4. Who talks to whom?
5. Who is in charge?

The various links that connect the members of such a team model themselves can be studied to measure the rate, volume, and costs of intrateam transactions. The model's performance with respect to its task environment serves as the primary measure of the modeled organization's effectiveness.

Extensive and detailed experiments with human organizations are time-consuming, expensive, and notoriously difficult to control, when they are feasible at all.

Heuristic models of organizations can be expected to help overcome each of these obstacles by opening up new horizons in organizational engineering. Authority, responsibilities, resources, task definitions, and communications structures can be freely manipulated to a degree quite unthinkable in traditional organizational research. It is unlikely that model-based experimentation of this kind could (or should) become the sole basis for the design of human institutions. Pather, a model-based approach to organizational engineering will permit and encourage the rapid and economical exploration of a broadened range of social hypotheses. The results of such work, having been obtained at low cost to experimenter and subject alike, will serve to focus confirming human experiments on the most promising alternatives.

SOME RESEARCH AND DEVELOPMENT THEMES

Now that we know what heuristic models are, how they are built, and what important benefits they offer in the field of organizational design and training, I would like to conclude this paper with some general remarks aimed at research and development sponsors: How to spend funds to encourage the realization of these benefits.

The first and most urgent need is increased support for the *art* of model building. Cognitive psychologists must be encouraged in their efforts to uncover how people acquire, process, and transmit information. Artificial intelligence experts need to be supported in building better modeling tools. On the one hand, better languages and systems—like ROSIE—are needed; on the other, automated model building through exemplary programming must be brought to maturity. Much of this work can be effectively driven through energetic support for model-building applications; such applications not only gauge the state of the art, they also serve as a powerful stimulus for tool building and cognitive research.

In addition, research sponsors should actively assist and encourage much greater contact between model builders and human experts in the model domain; access to real experts is never easy, and it has in some critical instances proven to be nearly impossible.

For the time being, the *engineering* aspects of model building are reasonably well served by the continuing surge in hardware design and fabrication. The availability of fast, low-cost, powerful, and compact computing systems assists model builders as much as the rest of the population. The same is largely true for software engineers: Better algorithms, more reliable codes, and improved man-machine interfaces are topics of wide concern throughout the computing community.

The *industry* of heuristic model building—moving models from the laboratory into the workplace—will require considerable support over the next decade from R&D sponsors. The problem here, *technology transfer*, and its solution will require carefully integrated efforts from the research and application communities. The potential impact of heuristic modeling on the Department of Defense, civilian government organizations, the educational establishment, and business and industry is very great. Today, however, none of these communities is well equipped to make use of this emerging technology. Possibly the most effective device for encouraging the large-scale organizational initiatives that will be required is the funding of carefully planned, generously supported, highly visible *showcase* projects—aimed to attract maximum attention and to incite emulation.

DISCUSSION OF SESSION 6

Rick Hayes-Roth began the discussion by comparing Nilsson's work to current research on semiautomated tank networks. He pointed out a natural extension of Nilsson's research from totally automated teams to teams in which some participants are human and others are intelligent machines. He also suggested exploring how the relative intelligence of team members affects performance by varying the "intelligence" of robot team components.

A dialogue ensued contrasting heuristic models of expert knowledge with mathematical policy models. Edwards noted that research using linear models to simulate expert judgment assumes that such judgments are inherently probabilistic. Programs such as MYCIN, which also purport to model expert decisions, appear to be inherently deterministic. He asked for a resolution to this conflict. Martins replied that simulations vary in depth but that all include a probabilistic element at some level. In linear models, the "dice-throwing" occurs at a very shallow level, whereas in heuristic simulations, it is buried deeper in the details of the model. Edwards asked why deeper analysis is required if our interest is in effective performance. Martins noted two reasons why a process model might be superior to a more superficial mathematical model. First, the trace of inferences used to arrive at a decision can serve as an explanation for that decision. This can be useful, since experts are often required to justify their decisions. Second, the model provides a causal framework for understanding the effects of various actions or constraints on performance. One can alter the internal processes, observe the results, and hopefully learn more about the processes that generated them. Martins acknowledged that heuristic models face a difficult challenge in producing performance comparable to that of linear models.

Rick Hayes-Roth argued that it is erroneous to associate heuristic models with deterministic operations. MYCIN and HEARSAY have built-in uncertainty factors, although formal statistics play a small role in their operations. Edwards objected to the implication that expert judgment is deterministic. However, Martins maintained that this is not an assumption of heuristic modeling, but rather that randomness is less obvious in a heuristic model because such models describe decision processes in greater detail than do linear models.

Thorndyke changed the topic by asking Nilsson whether studying communication problems in real-world teams will ever be feasible, given the difficulty of the problem in the simplified world of robot teams. Nilsson argued that once solutions were found for the hardest problems, extrapolation to other environments would be relatively easy. Hunt was skeptical, pointing to the trend in artificial intelligence toward programs with a great deal of specific knowledge about narrow domains. However, Nilsson maintained that certain general and fundamental problems (e.g., communication and inference) must be addressed in any sort of team and in any environment. Rick Hayes-Roth pointed out the constant tension in AI research between a high-level, synthetic approach and a more specific approach dedicated to case studies of real-world domains. Several participants concluded that both approaches can be useful and that each can benefit from progress in the other area.

VIII. DECISION THEORY

IN DECISIONMAKING, ARE N HEADS BETTER THAN ONE?

Ward Edwards

The classical theory of decisionmaking assumes that the function of any decision is to maximize the utility or expected utility of the decisionmaker, who is assumed to be rational. The unreality of the assumption that men are always rational has received almost unlimited attention—indeed, in my view far too much. Although I would never defend the economist's notion of a Rational Man, I do defend the idea that real men do a pretty good job, given enough time, resources, and reason to work at the problem. I consider it a gross exaggeration to label men as "intellectual cripples," as a recent discussion of cognitive biases did. However, that is not the point of this paper.

No one who has participated in a structured organizational decision process can accept the idea that there is a single decisionmaker whose utility is to be maximized. This presents the formal theory with an insoluble dilemma, since conflicting interests seem to make the notion of maximization useless—and without that notion, or an ideal closely related to it, we do not know how to give a sensible account of actual decisions, and we certainly do not know how to improve them.

There is indeed a Decisionmaker—The Boss. The Boss serves a wide and important variety of functions, but they do not include decisionmaking. One function he obviously serves is to "sign off" on the decision. This may mean two things: He can veto a decision that does not make sense to him, and he will take the blame if things go wrong. Of these two functions, the latter is probably the more important and the more often exercised.

In deference to the label of this conference, I will call those who support The Boss in decisionmaking his Team. The Boss's Team is typically, for important command decisions, composed of experts; usually each knows more than The Boss does about some topic involved in the decision problem. Moreover, the Team has much more time, both as individuals and in meetings, to consider the problem at hand, formulate options, collect the information relevant to evaluating them, perform the necessary massaging and packaging, and come up with a recommendation and a justification for it, all in neatly structured form. Depending on the individual style of The Boss, this process can go on almost without his input until the very end, or he can be close to every detail. There is some research evidence and some experience to suggest that, of those two extremes, the former is preferable.

The Boss also is an expert. Typically, he is expert about the relations between the organization he is Boss of and other external organizations—often those higher in the command hierarchy. On such issues as the constraints imposed from above and the values at higher levels with which the decision should be concerned, The Boss is likely to have the most definitive available information. Thus The Boss, if he is good at his job, performs both as leader of his Team and as a skilled Team member.

I do not believe that the fact that a decision made by a Boss and his Team, working together, makes the idealization called maximization of utility, or of ex-

pected utility, inapplicable to real organizational decisions. The idea that makes decision rule work is the idea of an organizational utility function. It is reasonable to suppose that organizations, like individuals, have values. Such values will ordinarily be multiattribute, and different members of the organizations will be expert about different attributes.

This idea appears in most decision-analytic practice, especially in military contexts. Seldom are all judgments required for a decision analysis made by one person. Such judgments are usually of either three or four kinds. One concerns the structure of the problem—the options, the decision tree (if any), the attributes of the multiattribute utility functions. Structuring a problem is typically an all-hands task in which The Boss and his Team all participate. The second concerns measures of single-dimension utilities—the locations of the objects being evaluated on the various dimensions of evaluation. These are typically judgments requiring physical measures, expertise, or, more frequently, both. The Boss will not ordinarily want to make such judgments except in areas of his special expertise. The third concerns weights. Weights are, in my view, the essence of value judgments. As such, they may be made by The Boss, but they are seldom made alone. Finally, for problems in which uncertainty is important, judgments of probability may be needed; again, these require special kinds of expertise.

So far I have discussed decisions only at a single level within an organization. Quite often, important decisions span several levels. In that case, it is possible and valuable to construct a hierarchical value structure isomorphic with the hierarchical command structure of the organization. Thus the value labels and weights appropriate to a given command level can come from that level. As usual, location measures and probabilities will normally come from appropriate experts. This idea of value hierarchies isomorphic with command hierarchies is much newer than the idea of organizational values; I know of only one or two applications. I expect to see more as the idea spreads and the technology develops.

Finally, the problem gets still more complex when the organization cannot be regarded as the decisionmaker. For most really important decisions, there not only is no single individual but also no single organization that serves as decisionmaker. Instead, there are a number of organizations with stakes in the decision; typically no one has command authority over any other. A decision affecting more than one service made at the JCS level is an obvious example, though less exalted ones are easy to find—e.g., in any joint operations context.

I speculate, and have a few instances to back the speculation, that the values of conflicting stakeholder groups are similar in structure but differ in weights. If so, one approach to such decisions is to use the structure as a heuristic device to facilitate bargaining and negotiation; this technique has been used in several international negotiations. If a higher-level decisionmaker exists, a much more radical idea is to have that decisionmaker weight the stakeholder groups; this would permit aggregation of values across such groups.

Seaver (1978) recently reviewed the literature on probability estimation by real and imaginary groups (imaginary groups being obtained by averaging individual judgments). The review concludes that the evidence does not justify the hypothesis that n heads are any better than one. That is, averaged individual judgments are just as good as group judgments. However, almost all the studies used college students or other nonexperts as subjects, and none used groups as a means of combining different forms of expertise.

This introduces two of the research problems that seem to me to be latent in the view of decision-analytic work on group decisionmaking that I reviewed above. First, what is an expert? The topic is almost unstudied—mostly because the time and expertise of experts are too valuable to spend on research. Further, we do not have any systematic information about what happens when substantive experts disagree in an area of their shared expertise, or what happens when an expert disagrees with someone who has an amateur's knowledge of the topic of disagreement, while being himself expert on something else. We all have experienced such interactions, but I know of little technical to say about them.

Second, if a number of experts of different kinds must contribute to the solution of a problem, does it pay to bring them together around a table so that each in effect can spend some time teaching the rest about his or her field? Or is it better to find ways to partition the task so that each can work alone, leaving the aggregation of their individual efforts either to a formal aggregation rule or to a super-expert? Since both styles are likely to be attractive, a better formulation might be: What rules should govern the choice between these styles and how well are these rules applied?

I am fascinated by the role of expertise in bargaining and negotiation, in real situations that are not zero-sum. In part, the expert may help to identify the solutions on the Pareto frontier. Beyond that, he is likely as a committed member of his Team to try to promote that Team's welfare. These two roles seem to conflict—yet we see resolutions of them without bloodshed in every newspaper. Indeed, this combination of disinterested expertise and highly interested advocacy characterizes most real-world negotiations. We have not done very well at bringing the process I have just described into the laboratory.

Finally, I am so charmed with the idea of negotiating over values instead of over decisions that I would like to see some study of whether that is ever possible. As far as I know, it has never occurred in a real negotiation. Yet it might be incredibly efficient compared with what we do now.

STUDYING TEAM PERFORMANCE: A DECISION THEORY PERSPECTIVE

Kenneth R. MacCrimmon

DECISION THEORY: A QUICK OVERVIEW

The term "decision theory" is used in various ways. On the one hand, the scope may be viewed so broadly as to encompass virtually any theorizing about thought and action related to human choice behavior. The origins of this approach can be traced back at least as far as Aristotle and his focus on "proairesis" or "deliberate choice" (see *Ethics*, Book 3). In a narrower sense, decision theory is often used to describe that research in economics, psychology, and related disciplines that deals with individuals making utility-maximizing decisions. Even this research spans a broad spectrum; in these few pages available for taking an overview of decision theory we shall concentrate on the research emphasizing utility-maximizing choices. The omission of the literature on information-processing limitations, irrationalities, and so forth is unfortunate and will to some extent be remedied in the final section of this paper.

The decision theory—*qua* utility-maximization—research has traditionally had an individualistic focus. Thus when groups or societies have been studied they are assumed to be composed of sets of utility-maximizing individuals, e.g., in a market context. At times, though, a wholistic focus is taken and one talks about the goals of a business firm or the beliefs of a nation.

In this section we shall adopt an incremental approach. First we consider a single individual making choices that affect only himself. Then we introduce another person, not as a separate decisionmaker, but solely as a passive recipient of the outcomes of the first person's choices. Hence the decisionmaker may wish to take into account the effect of his choices on this other person. Next we shift to a real two-person team. Each team member can make choices and receive outcomes. Finally, we examine the effect of adding a third, fourth, and eventually an *n*th person.

To provide an overview within just a few pages means that explanations of concepts, theories, etc., cannot be given. We shall assume that the reader is somewhat familiar with this literature. Our task then will be to provide an organization of some of the important research areas that may help to jog the reader's memory or stimulate the reader's thoughts. To someone without any knowledge of decision theory research, the material in this section will probably be a meaningless jumble of terms. With that caveat, let us proceed to take a quick look at work in decision theory that could be relevant to a study of team performance.

To the extent that we view a team as composed of a set of individuals, it is important to develop a careful model of individual behavior before jumping to a complex team. Table 1 summarizes some key topics in single-person choice. As the table suggests, the bases of individual performance are the beliefs, values, and resources of the individual. Belief and value assumptions form the central part of any decision theory. Resources include both the internal capabilities of the individ-

Table 1

KEY CONCEPTS FOR THE SINGLE PERSON TEAM

BASIC ELEMENTS: Beliefs/Values/Resources/ Disposition: mode of rationality	
CERTAINTY	UNCERTAINTY
Single Objective: ordering, consistency, maximize utility, gains and losses, net values, available alternatives, feasible set, inputs-outputs, production function Multiple Objective constraints: conjunctive-disjunctive, dominance, sequential elimination, goal hierarchy, instrumentality, independence, weights, lexicography, interdependence, trade-offs	Information Collection: uncontrollability, state variables, EVPI, cost, Bayesian revision Choice: maximum expected utility, decision variables, subjective probability, utility functions, risk propensity, choice axioms, belief and value biases

ual and any externally held possessions or opportunities. The individual's disposition or mode of rationality could be considered as a part of the resources. A decision form of explanation would take the form: (1) The individual believed he was in a situation of type S (i.e., the belief premise); (2) the individual wanted to attain goal G (i.e., the value premise); (3) the individual strove to attain goal G in manner R (i.e., the resource/disposition premise); then (4) an individual pursuing goal G in manner R will, in situations perceived as type S, take behaviors X (i.e., the law premise). This explanatory form (an adaptation of Hempel's covering-law approach) is implicit in all the theories we shall mention.

The primary distinction we make in examining single-person behavior is in whether the environment is treated as certain or uncertain. There are a variety of theories which assume a certain environment so that attention can be focused on other issues. The theory of consumer demand has demonstrated the power of such simple models. Recently, considerable study has been made of individuals with multiple objectives trying to make value-maximizing choices over alternatives that contribute differentially to these objectives. When explicit account is taken of the uncertainty, generally by probability assessments, there are two main types of decisions: (1) choosing among the terminal alternatives (which may be sets of sequential actions), or (2) choosing among strategies for collecting more information before selecting a terminal alternative. The main underlying theory that emerges in this case is the maximizing of expected utility; that is, subjective probabilities are assigned to the state (or uncontrollable) variables, utilities are assigned to the consequences, and the alternative (perhaps involving information

gathering) yielding the highest expected utility is chosen. Associated topics involve the systematic revision of beliefs and the characterization of risk-taking propensity. In many multiple-person theories, one begins by assuming that each individual maximizes expected utility, hence the individual level deserves careful attention even when the concern is with team performance.

After a single-person "team," the next obvious step is to add a second person. However, before adding a full team member it is instructive to consider the case where the second person receives only outcomes based on the choices made by the first person. Thus the second person plays a passive role being unable to affect the choices directly. This intermediate case is interesting because it raises the question of the relative impact of the choices taken on both parties, without also introducing the complication of interacting choices. The emphasis here is strong on the values (as contrasted with beliefs) and various theories deal with how the choosing individual takes the interest of the other person into account both in a domain of ordinary choices and in an ethical domain. Table 2 lists some of the key concepts.

Table 2
KEY CONCEPTS FOR THE ACTIVE-MEMBER/PASSIVE-MEMBER TEAM
SINGLE PERSON AND ANOTHER

RELATIVE IMPACT	RELATIVE VALUES
Pareto frontier, efficiency distributional externalities, transfer payments, compensation comparability	Self-Assessment: self-interest, altruism, hostility, cooperation, competitiveness, relative contributions Ethical Concerns: utilitarianism, maximum distributional weights interpersonal comparisons other's shoes equity: inputs-outputs

Next we assume that the second person also has control over the outcomes, and so we come to a meaningful use of the term "team." The central focus is on the actions selected by each person, the payoffs attained, and the distribution of the payoffs. The primary distinction is on the degree of commonality of interest (i.e., values). In Table 3 we consider the extreme cases of no commonality and complete commonality. While a set of two individuals with directly opposed interests would probably not be called a "team," it does represent a special case of considerable interest due to the game theory results pertaining to zero-sum games. The case of completely congruent interests certainly does deserve the label "team," and even

Table 3

KEY CONCEPTS FOR THE TWO-PERSON TEAM (EXTREME INTERESTS)

<u>DIRECTLY OPPOSED INTERESTS</u>	<u>COMPLETELY CONGRUENT INTERESTS</u>
relative control over actions/ payoffs, normal-extensive representation, common unit of value, zero-sum, maximum	Communication: content: alternatives, beliefs, objectives, intentions channels: capacity, rate, reliability timing: sequential, simultaneous coordination, costs, information structure, differential info, sharing, expertise
<u>COMPLETELY CONGRUENT INTERESTS</u>	
No Communication: implicit coordination, unilateral, mutual adjustment	

though it is an extreme assumption, it does highlight issues of communication and coordination without confusing the study with value differences. This case includes the formal theory of teams discussed in the second section of this paper.

Table 4 extends the two-person team to the general case of mixed interests. As the listing suggests, there are theories in virtually all the social sciences that relate to this situation. Communication possibilities are again an important consideration. Differential skills and resources become a primary focus especially with respect to the differences in values and beliefs of the team members. Even though

Table 4

KEY CONCEPTS FOR THE TWO PERSON TEAM (MIXED INTERESTS)

<u>MIXED INTERESTS</u>	<u>DIFFERENTIAL RESOURCES</u>
No Communication: non-zero transferability, individual vs. group rationality, knowledge of actions, payoffs, problem types: prisoner dilemma, etc., solution possibilities, metagame, equilibrium, stability, randomization, mutual expectations	task differences, allocation, resource specialization, flexibility, different levels, different skills, division of labor, comparative advantage, resource match, synergy, role differences, leadership, agency control over decision variables, payoffs, risk sharing, moral hazard, enforceability, con- tracts, rules, penalties
Communication: identification of common interests, gains from exchange, market, bargaining: offers, threats, concessions, social exchange	

a team of two members seems quite restrictive, we can see that many interesting issues can be addressed before moving to a more complex model.

Clearly, we want to advance beyond the two-person case because even adding a third person introduces two versus one issues. While it would be reasonable to first consider a third team member (perhaps initially as a passive agent and then in an active role), we shall jump to an n -person case—where n can be 3 or some much larger number. A useful distinction in the n -person case is whether one assumes that the individuals each can take individual actions that mutually determine the individual and joint outcomes or whether the individuals are viewed as voters trying to influence the overall action of the collective. Table 5 presents some relevant topics on these issues along with further attention to the differential resource case, such as rule hierarchies.

In a few pages, obviously one cannot do more than skim over the surface of some very involved theories and models. It should be apparent, however, that there are many results that are already of relevance to team performance and that a number of promising directions remain unexplored.

Table 5
KEY CONCEPTS FOR THE N -PERSON TEAM

<u>MULTIPLE INDIVIDUAL ACTIONS</u>	<u>JOINT ACTION</u>
Pareto frontier, size of pie, characteristic values, payoff dimensions, coalition possibilities, overlap, stable coalitions, core, counter offers, share of dividends, side-payments, ganging up, common enemy, markets, size, minimal winning, incremental contribution	aggregation of individual preferences, social choice axioms, voting procedures: unanimity, majority, sequential, agenda, log-rolling, tyranny of minority, majority regulations, rule-setting.
	<u>DIFFERENTIAL RESOURCES</u>
	relative power, contributions, resource control, intervenor, facilitator, exploiter, hierarchy, role differences, incentives

FORMAL THEORY OF TEAMS

In extending individual decision theory to teams, the natural focus is on multiple individuals sharing the same utility function and having to make decisions in an uncertain environment. By allowing team members different opportunities for observation (of the environment) and communication (with other team members), we can ask for the characteristics of the information-decision structure that will lead to the best team performance. This is the model used by Marschak and Radner (1972) in their economic theory of teams. Even though a number of restrictive assumptions have to be made for mathematical tractability, the theory provides a clear benchmark for optimal team design. When we study how people actually

behave in situations for which we can compute the optimal performance, we may gain some insight into the mistakes and fallacies that occur in real teams operating in very complex environments.

The basic notation of team theory is that of expected utility theory. Let $X = |x|$ be the set of (uncontrollable) states of the world and $A = |a|$ be the set of actions available to the decision-maker. If $\omega(\cdot, \cdot)$ mapping from $X \times A$ to the real line is the utility function, and $\pi(\cdot)$ mapping from X to $[0,1]$ is the (personal) probability function, then an individual's expected utility for action a is

$$\sum_x \omega(x, a) \pi(x).$$

In a multiple-person situation, where the actions available to person i are $A_i = |a_i|$, a is the joint action

$$a = (a_1, a_2, \dots, a_n) \in A_1 \times A_2 \times \dots \times A_n.$$

As the notation indicates, it is assumed that all members have the same utility function and the same probability function. As in the single-person case, the problem is to choose that set of actions that will maximize expected utility.

At a less aggregate level, it is clear that the actions of a decisionmaker should be based on the information available. Let the message set of person i be $Y_i = |y_i|$ and the decision function or strategy of person i specifying particular actions to be taken in response to particular information be $\alpha_i(\cdot)$, that is $a_i = \alpha_i(y_i)$. The next step is one of relating the information received to the states of the world. Let $\eta_i(\cdot)$ be a measurable function from X to Y_i , i.e., $y_i = \eta_i(x)$, and call it the observation or information function. The team's expected utility of using an information structure $\eta = (\eta_1, \eta_2, \dots, \eta_n)$ and a team decision function $a = (a_1, a_2, \dots, a_n)$ is

$$\sum_x \omega(x, \alpha_1(\eta_1(x)), \alpha_2(\eta_2(x)), \dots, \alpha_n(\eta_n(x))) \pi(x).$$

Call this expression $\Omega(\eta, a; \omega, \pi)$, since it is a function of η, a, ω , and π . It is assumed that the utility and probability functions used are the true ones, thus they are not decision variables, as are η and a ; and hence, for simplicity, we shall write $\Omega(\eta, a)$ in place of $\Omega(\eta, a; \omega, \pi)$. The problem then becomes one of choosing, for a given information structure η' , the best decision rules, that is, the $a = a^*$ for which

$$V(\eta') = \max_a \Omega(\eta', a).$$

After each information structure is evaluated in terms of the best associated decision function, one chooses the information structure η' that maximizes $V(\eta)$.

Rather than presenting a formal treatment of team theory, we shall describe a particular example. In MacCrimmon (1974), a team problem involving three hypothetical sugar plantations in the Philippines is presented. Figure 1 presents a complete problem description. The team of the three plantation managers must make decisions in the face of uncertainty about their costs. The key questions to answer are (1) What observations should be made on the environment (i.e., which plantation managers should purchase cost forecasts)?, (2) Who should communicate with whom (i.e., what messages should be sent between plantation managers)?, and (3) What decisions should be made (i.e., which plantations should harvest)?

This problem, although quite simple, allows many of the features of team theory to be illustrated. The non-smooth payoffs, on the other hand, do not allow

Team Theory Problem

Introduction:

You have been called in as a consultant to the Sweetsa Plantation, a multi-site plantation system which grows and exports sugar in the Philippines. You are presented with the accompanying information and are asked to make some suggestions as to how the plantation system is to be organized and run.

Problem Description:

The company has three geographically separate plantations of varying efficiency. (These will be designated as the East, West, and South plantations, to correspond to their geographic positions.) Each plantation is capable of producing 1,000,000 pounds of sugar each quarter, i.e., each three months. Each quarter represents a different harvest subject to different weather conditions.

Each plantation has a fixed quarterly cost of \$15,000 due to the costs of land preparation, planting, land maintenance, capital and administrative costs, etc. At the beginning of the quarter, the manager of each plantation must decide whether or not the current crop (of 1,000,000 lbs.) is to be harvested, an action which incurs the additional costs of the harvesting, processing and shipping. Because of the uncertainty about weather conditions throughout the quarter, labor costs, transportation availability, etc., the expected costs may be so high that it might be more profitable to simply let the sugar go unharvested and simply to be plowed under for the next quarter's preparation. As noted above, these variable costs vary widely: on the Eastern plantation, the variable costs for harvesting the crop are \$40,000 if good conditions prevail, \$150,000 if bad conditions prevail. On the Western plantation, the variable costs are \$60,000 for good conditions, \$140,000 for bad; on the Southern plantation the variable costs are \$70,000 for good conditions, \$140,000 for bad; on the Southern plantation the variable costs are \$70,000 for good conditions, and \$130,000 for bad conditions. The cost history for the past ten quarters is as follows.

Plantation	Quarter									
	21	22	23	24	25	26	27	28	29	30
East	150	150	40	150	40	150	40	40	40	150
West	60	140	140	60	140	60	60	140	140	60
South	70	130	130	70	130	70	70	130	130	70

(Figures are in thousands of dollars.)

The market for the sugar is in the United States, where various licensing and import regulations, tariffs and the like have created the following market conditions:

- 1) The Sweetsa Plantation can always sell the first 1,000,000 lbs. at 11c/lb. (\$110,000).
- 2) The second 1,000,000 lbs. will only bring 10c/lb. (\$100,000), while
- 3) The third 1,000,000 lbs. will bring only 6c/lb. (\$60,000).

To aid the plantation managers in the making of their decisions as to whether or not to harvest their crops and incur the variable costs, each manager has the opportunity to commission an extensive study of his quarterly conditions, including weather forecasts, labor supply determinations, transportation studies, etc. This study will provide him with prior information about whether the variable costs on his plantation are to be high or low for that quarter. It can be assumed that this study provides perfect information, i.e., it predicts perfectly all the time. The cost of conducting such a study is \$12,000 per quarter for each manager who decides to take advantage of it.

Furthermore, the plantations may set up a communication system which allows a manager to convey his own cost-information (as determined by his commissioned study) to any other plantation manager. Each such one-way communication between plantation managers essentially costs \$2,000, the cost of providing the communications links. So, for example, East can communicate to West and the Sweetsa Plantation incurs a cost of \$2,000 for the communication link and a cost of \$12,000 to obtain East's data. The end result is that both East and West know what the cost conditions are for East during that quarter.

Consultant's Role:

Given all of the information above, Sweetsa has asked you to develop an optimal organizational structure and its concurrent set of decision strategies. That is to say, you are to specify which plantation managers are to commission the variable-cost studies, who shall communicate to whom, and what decisions are to be made with reference to the sugar harvesting.

Fig. 1—The sugar plantation problem

one to apply the usual team theory theorems or algorithms. The computation of the optimal solution and the evaluation of the nonoptimal solutions are made in terms of straightforward expected-value calculations. There are enough twists, however, so that the optimal solution is not immediately apparent. The use of seemingly reasonable heuristics, such as observing the most variable part of the environment, lead one in the wrong direction. This allows the problem to be used in empirical studies directed toward understanding team performance.

Table 6 presents a summary of the results obtained in an experiment with 60 upper-level managers. The optimal solution, comprising one observation (by West) and two communications (to East and South), was recommended by only 6 managers (i.e., 10 percent). The interesting issues arise from examining the nature of the solutions recommended by the remaining 90 percent. As the figure shows, over one-third of the subjects made infeasible recommendations. The most common fallacy was to expect a team member to act on information that was available within the team but had not been communicated to that person. Even among the feasible rules, we see a tendency to waste or misuse observations and communications. The phenomena that occurred in this simple situation seem representative of what happens in real teams.

The results can be examined in terms of the information available to team members. Note first the cases that involve too little information within the team. There may be too few observations or too few communications. In one instance, information potentially available could have increased gross profit more than the increase in cost—but it was not collected or distributed. Another instance involves

Table 6

SUMMARY OF RESULTS IN SUGAR PLANTATION EXPERIMENT

Characterization of Solution	Number of Subjects
Feasible Rules:	38
The optimal solution	6
The four next best solutions	12
Shut-down of all plantations	6
Wasted observations or communications	4
Misused observations or communications	5
Poor observations or communications	3
Apparent transcription mistakes	2
Infeasible Rules:	22*
Communicating data not observed	2
Acting on data not observed	7
Acting on data not communicated (or observed)	5
Acting on data not communicated (but observed)	14

*Note: Some subjects had several infeasible rules.

information which is expected by particular team members but is not made available.

Correspondingly, there were several cases involving too much information within the team. In one case, information collected or distributed when used fails to increase expected gross profit by as much as its cost. In another case, the benefits of information not only do not exceed the cost but the use of the information actually decreases expected gross profit. This happens due to interference of one decisionmaker with another, i.e., the "too many cooks" syndrome. In other cases, after the cost and effort of collecting and distributing information, the information is not used.

The economic theory of teams, then, not only represents a formal mathematical prescription of team behavior, it also serves as a model for studying information processing as it occurs in teams. The example described here, and the particular questions asked, just begin to tap the possibilities for using this theory. For an indication of some of the other research directions see MacCrimmon (1974).

STUDYING DECISIONMAKING TEAMS: ELEMENTS OF A RESEARCH STRATEGY

Underlying the team problem discussed in the preceding section are the same three elements (beliefs, values, and resources) highlighted at the beginning of this paper. These concepts are the important characteristics in any study of humans operating in real teams. In MacCrimmon (1970), beliefs, values, and resources are used as a unifying framework in the process of drawing propositions from major descriptive studies of decisionmaking.

Assumptions about beliefs, values, and resources provide a useful way to build models of team performance. By starting with the simplest kind of team, we can extend the model in the desired directions until we obtain the type of structure needed for the questions being addressed.

Our starting point is a "pure team," which we define as a collection of individuals having identical beliefs, identical values, and identical resources. Each of these individuals is assumed to be value-maximizing, subject to their knowledge and processing abilities. They take actions that yield individual and joint outcomes. This is represented by the center column of Fig. 2 which will serve as the focus for the following discussion.

First, we allow for different beliefs. It is useful to distinguish between individuals who have different beliefs because they have different information and those who, due to different belief functions, would have different beliefs even with the same information. The first viewpoint leads us in the direction of communication structures and information sharing. This is the key problem addressed in the economic theory of teams. It also seems to be the basis for techniques of expertise information flows, such as Delphi. Both viewpoints suggest the concept of Bayesian teams. The second viewpoint involves different perceptual maps.

Second, we modify the assumption of identical values. The first modification is to assume a different utility function within each attribute of interest. This suggests differences in risk-taking propensity. Then one can look at differences in interattribute weights. Such differences imply variations across individuals in how they

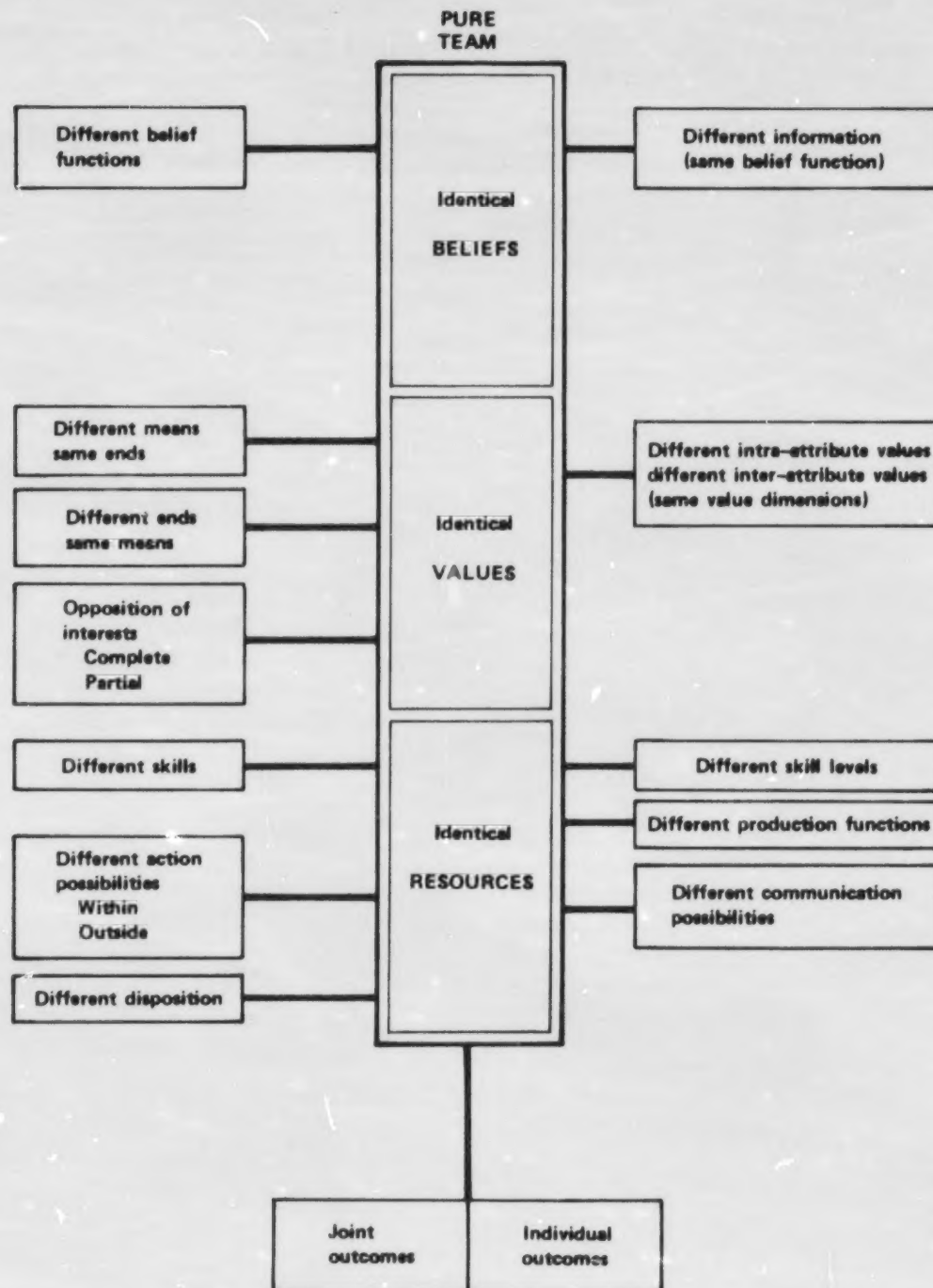


Fig. 2—The "pure team" and its variants

judge priorities, importance, or salience. In each of these cases, it is necessary to recognize the zones of agreement (either on means or on ends) among the participants. The ease with which these commonalities can be identified and pursued has major implications for team performance.

Finally, we remove the assumption of identical resources. The most minor change is to allow for the same types of skills but different skill levels. Then we allow for different types of skills. This raises the questions of comparative advantage, specialization, and division of labor. We also want to examine the effect of the relationship between each individual's inputs and outputs, i.e., the production function. Since communication is such a central topic in teams, we need to investigate the result of different communication possibilities. A natural follow-up is the study of different action possibilities both inside and outside the team.

By starting from the pure team model, we can make such incremental changes in a sequential or simultaneous fashion and trace through the effect on the behavior and outcomes of individual team members and the team as a whole. This approach is summarized in Fig. 2.

Two basic resource-related changes that are worth looking at in detail are the assumptions about the mode of rationality and about the specialization related to phases of the decision process. As noted earlier, there is a range of assumptions made about the capability and motivation for the team member to proceed rationally in thought and in action. At one extreme, theories may assume that an individual is comprehensively rational in that he has a formal, consistent system of beliefs, elaborate goal hierarchies, and sophisticated information processing capabilities. At the other extreme, the individual may be assumed to behave in a purely dispositional manner with stereotyped and inconsistent beliefs, vague and confusing values, and very limited means for thinking. Actions would be primarily based on habits and standard operating procedures. Between these extremes, one can make a variety of assumptions about limited rationality. A summary is given in Table 7.

Table 7 also shows the effect of the rationality assumptions on the decision process. While there are many different characterizations of decision process, we use pre-choice (called "observation phase"), choice, and post-choice (called "realization phase") as the distinguishing aspects. Any decision situation involves at least these three stages. A team member may participate in each of the phases or may specialize. In differentiated teams, we expect to see some members or units primarily serving an information gathering and processing role. Other units would specialize in performing the tasks necessary to realize the choice. Some units would have primary responsibility for making the choice. In studying decisionmaking in teams, it is useful to identify the type and degree of specialization of each member and the associated similarities and differences among members. One can use this process-based characterization of members' roles as a key aspect of guiding a study of team behavior. A diagrammatic representation of the relationship between beliefs-values-resources and the process/unit specialization is given in Fig. 3. For a further discussion developing a process approach see MacCrimmon (1973), in which the decisionmaking literature is interrelated in a process framework.

In addition to focusing on the members and internal workings of a team, it is desirable to characterize the environment facing a team. MacCrimmon (1970) discusses more than a dozen key characteristics. In MacCrimmon and Taylor (1976) this array is narrowed to three crucial features: complexity, uncertainty, and conflict. It should be apparent from the survey in the first section of this paper that

Table 7

RATIONALITY ASSUMPTIONS ON DECISION ELEMENTS AND PROCESS

	Comprehensive Rationality	Limited Rationality	Dispositional
	Decision Elements		
Beliefs	Systematic Consistent Comprehensive Formalized Probabilities Bayesian revision	Few, focal events Myopic Crude expectations Resistance to change	Stereotyped Biased Uncertainty avoidance
Values	Multiple objectives Goal hierarchy Clearly specified Utility form	Aspiration levels Sequentially considered Short linkages	Unexplicated Little motivational force
Resources	Unlimited capacity Complete knowledge	Limited memory Constrained processing Unaware of capabilities	Routine Very restricted
	Decision Process		
Observation process	Unbiased Fine discriminations Thorough search Economic assessments	Problem triggered Local search Incomplete Restricted	Biased perceptions No active search Remedial focus
Evaluation process	Complete options/outcomes Total consistency Clear criteria/preferences Optimize	Focus on major outcomes Few alternatives Short-run view Satisfice	Little evaluation Trial and error Ignorant of consequences
Realization process	Full implementation Control Incentives Contingency plans	Sequential Incremental Adjustments	SOPs Orientation to status quo Wait and see

each of these factors serves as a central focus for the major theories of decisionmaking. We shall not treat these aspects further here, but see MacCrimmon and Taylor for a discussion of about 100 substrategies for dealing with these various types of environments.

The team and environmental characteristics that have been identified can serve as a basis for building a series of collective decision models. If the characterizations are useful, each of the models should provide a framework for a somewhat different set of questions. A study of a real situation through the use of these multiple models should provide a much more complete understanding of team performance than any single model. The virtues of such an approach were well demonstrated by Graham Allison (1968) in his study of the Cuban missile crisis. What we are proposing (although sharing a multimodel viewpoint with Allison) is a much more sys-

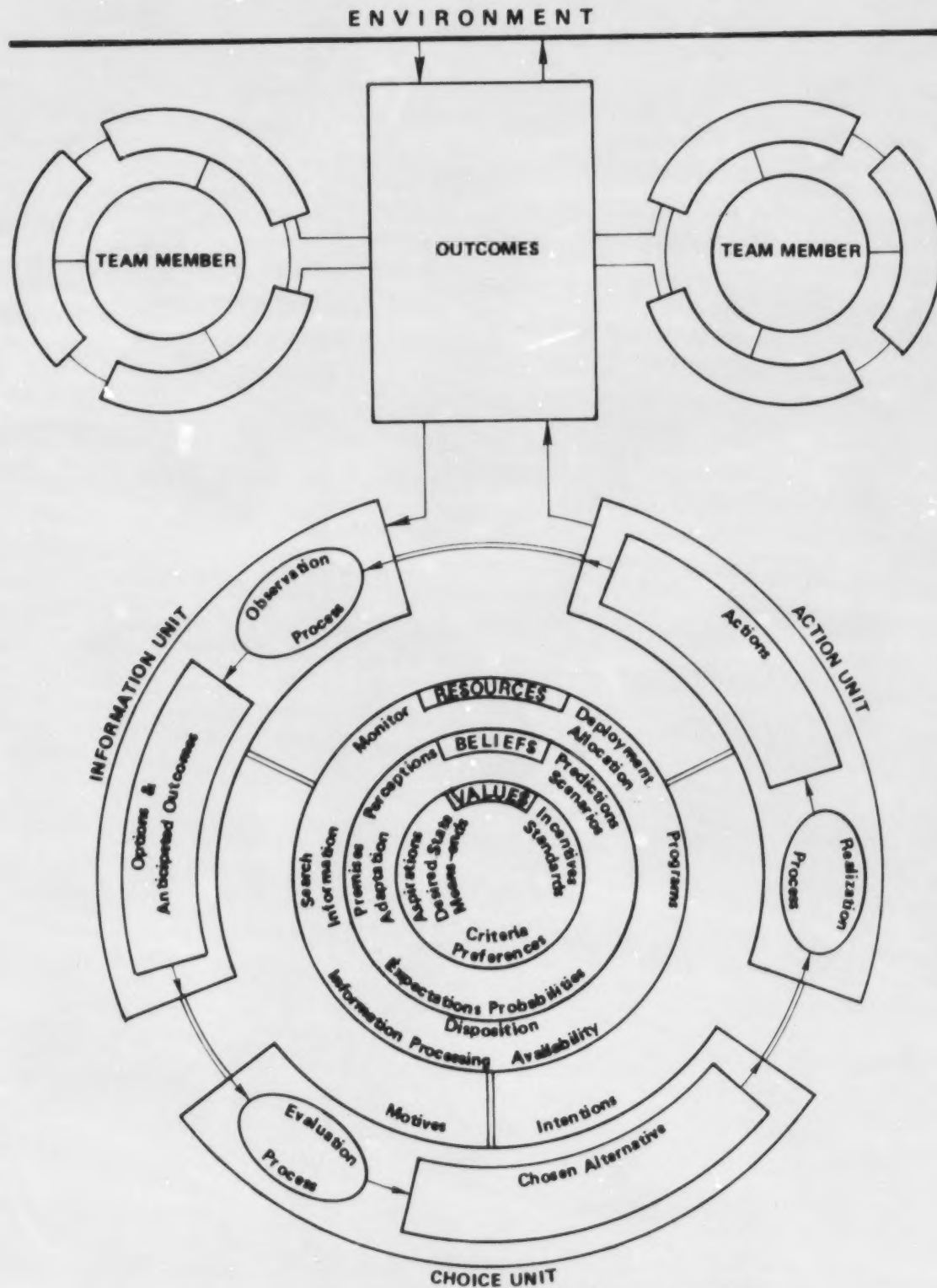


Fig. 3—Decision elements, process, and units

tematic approach with the underlying concepts, variables, and relationships clearly identified and with wide-ranging coverage of the relevant social science literatures. Some early directions were indicated in MacCrimmon (1973b). Initial applications have been made to the collective decisions involved in (a) dropping the atomic bomb on Hiroshima, (b) attempting to avert the bankruptcy of Rolls-Royce, (c) handling the FLQ crisis in Quebec, and (d) dealing with the Munich Olympic terrorism. Table 8 pulls together some of the characteristics we have been discussing and gives an example of some of the questions that are suggested by the various models. Further developments in this direction offer the prospect of utilizing decision theory (viewed broadly) in increasing our understanding of the behaviors of teams and other social units.

Table 8

COLLECTIVE DECISION MODELS AND ASSOCIATED FOCAL ISSUES

MODE OF RATIONALITY TYPE OF ENVIRONMENT		UNIFORM TEAM (Common Resources)		SPECIALIZED INFORMATION AND ACTION UNITS			
		CONGRUENT BELIEFS AND VALUES		CONGRUENT BELIEFS AND VALUES		DIFFERENT BELIEFS AND VALUES	
		COMPREHENSIVE	LIMITED	COMPREHENSIVE	LIMITED	COMPREHENSIVE	LIMITED
SINGLE CHOICE UNITS	COMPLEXITY	How is the problem modelled? What trade-offs?	How do processing limits affect choice?	What task specialization? How is coordination achieved?	How are tasks assigned? Any backups?	How do incentives promote joint action?	How effective are training programs?
	UNCERTAINTY	What info is collected? Basis for probability assignment and revision?	How is uncertainty avoided?	What observations? Who communicates to whom?	How is uncertainty absorbed?	How are information biases overcome?	Do separate actions aggravate uncertainty?
	HOSTILITY	What equilibrium solutions?	Basis for expectations of actions of others?	How are internal and external stability achieved?	How to avoid habits being taken advantage of?	Plans for replacing vulnerable points?	Areas of internal agreement?
MULTIPLE CHOICE UNITS		DIFFERENT BELIEFS AND VALUES					
	COMPLEXITY	How are multiple preferences aggregated?	How are fragmented models connected?	How does leadership emerge?	What are costs of inadequate coordination?	Whose model of reality is used?	Who makes which adjustments?
	UNCERTAINTY	How is info pooled? Correspondence of scenarios?	Incomplete vs inconsistent perceptions?	How are contingency plans established?	What effect of environmental change on SOPs?	Who controls info and action channels?	How is technology utilized?
	HOSTILITY	Which coalitions form? What payoff distribution?	How does common threat promote solidarity?	How to diffuse adaptive capability?	How to assure necessary actions are taken?	What relative power?	How do mindguards develop and operate?

DISCUSSION OF SESSION 7

Thorndyke questioned the applied payoffs of a decision theory approach to team decisionmaking. Edwards replied that decision theory could serve a prescriptive role by providing people with methods for structuring the problem. Recasting a dilemma in terms of options, values, and uncertainties can help improve the quality of the decision. MacCrimmon added that classical decision theory provides some indication of what variables might be important in affecting decisionmaking and how these factors might be related. However, he noted that it might be difficult to extrapolate from individual decisionmaking, to which classical models apply, to group decisions. It was also noted that current models are poorly developed in some areas. For example, little is known about situations in which information gathering and assessment have costs associated with them, as is frequently the case in military decisionmaking. Menchik noted optimistically that the formal theory of teams described by MacCrimmon focuses on many significant factors—skills, resources, goals, values, and so on—that might be further investigated in research on team decisionmaking.

IX. HUMAN ENGINEERING

HUMAN FACTORS IN TELECONFERENCING

Carl Feehrer

For the last couple of years, BBN has been working under contract with MIT/Lincoln Laboratory and the Defense Communications Agency (DCA) on an evaluation of different methods of speaker selection in limited capacity, secure-voice networks. Our experience during this evaluation suggests that the constraints placed on participants in conferences using such networks are different enough from those normally experienced in everyday face-to-face and short-distance telephone communications that it may be important to develop formal training paradigms for individuals and teams who will use future communication systems for teleconferencing. Unfortunately, at the moment, we are unable to say what the contents of those training programs should be. We are aware, however, of some of the phenomena associated with conversation over systems with certain speaker selection and delay characteristics, and we can at least begin to outline needs for research.

This paper will begin with a discussion of some of the difficulties that can arise in voice networks where conference participants communicating via satellite are separated from each other by different amounts of delay. Figure 1 depicts a situation where three conferees respond as quickly as possible to an input made by a fourth. From the latter's point of view, of course, the responses are distributed in time in accord with the length of the delay between him and each of the remaining participants. Note, however, that P2 and P3 collide—both are heard at the same time. This may or may not pose an impediment to P1's understanding of who spoke or what was said. Since we're presuming short "speeches" in this example, a conflict, if it occurred, could probably be resolved by simply asking for a repeat of the responses.

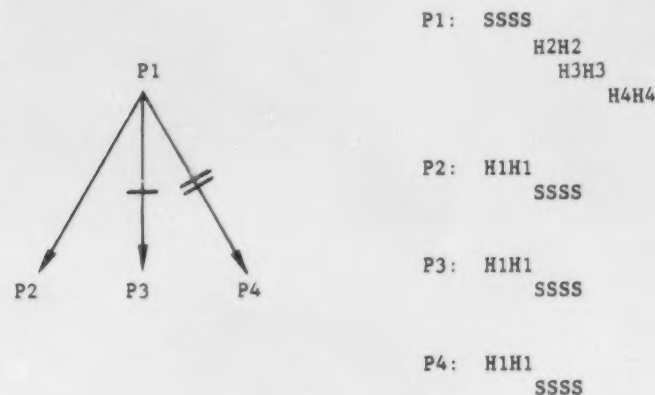


Fig. 1—Summary of what may be heard by P1 when three participants with different delay times respond as quickly as possible to an input by P1

Suppose, however, we are dealing with longer "speeches" and the attempts by one speaker to interrupt another. Figure 2 shows what may happen, even in face-to-face or other no-delay communications, and from this we can begin to guess what problems may occur if the situation is further complicated with delay. For later purposes let us define two major categories of interactive events—*interruptions* and *collisions*—and within each of these, two subtypes: for the *interrupt* category, "successful" and "unsuccessful" events; and for the *collision* category, "major" and "minor" events.

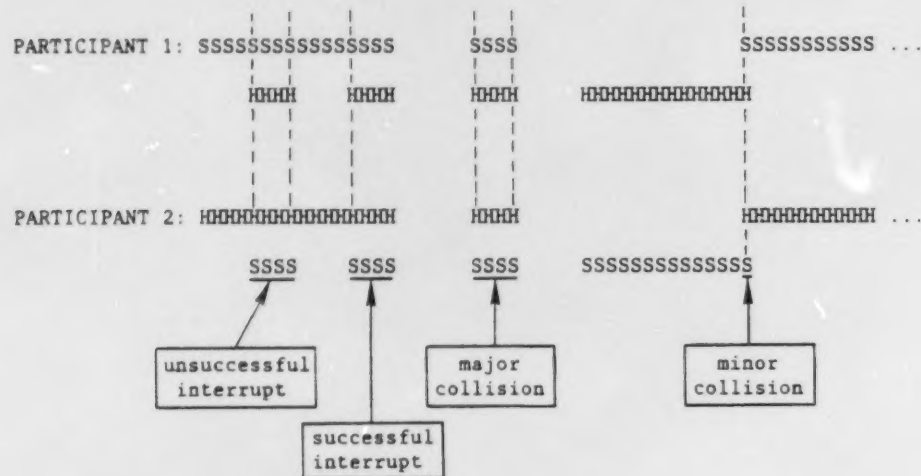
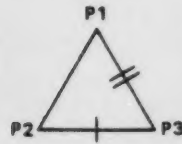


Fig. 2—Definition of interrupt and collision events

Now, let's look at the earlier network and focus on what P3 might hear if he were listening to a conversation between P1 and P2. At the bottom of Fig. 3 is a sample dialogue, fabricated with some gratuitous assumptions about how long it takes to say things. In this instance, the listener may not only *not* understand who is saying what, he may also not have been able to understand what it was that sank. And, if P1 and/or P2 continues to speak, it may be some time before he learns. Let's return to something more manageable than long speeches. Consider a very simple task in which each participant in a three-person conference has been given two words which he must speak as soon as possible after the participant before him has spoken. Figure 4 diagrams the first two cycles of such a task for three different types of delay networks: one (undelayed) telephone conferences, one with an even distribution of delays, and one with an uneven distribution of delays.

The primary things of interest here are the differences among the rhythms of speaking and hearing within and between delay distribution types. (For the distributions shown, rhythms are repeated every two cycles, that is, after each speaker in the triplet has spoken and been heard by the remaining listeners twice.)

Figure 5 summarizes the outputs from this very simple model of a communication task performed by three conferees with respect to number and type of collisions to be expected, "meter" (the rhythm associated with output of "speeches" in



P1: SSSSSSSSSS SSSS SSSS
H2H2 H2H2 H2H2

P2: SSSS SSSS SSSSSSSSSSSSSS
H1H1H1H1H1 H1H1 H1H1

P3: H1H1H1H1H1 H1H1 H1H1
H2H2 H2H2 H2H2H2H2H2H2H2

"in the clear"     

Speaker 1: The first ship that sank...first ship.....first ship.....

Speaker 2: Which ship? Which ship? I can't understand you, repeat.

Listener: The first.....sank... which...ip.. I can't understand.....repeat...

Fig. 3—Summary of speech heard "in the clear" by P3 during conversation between P1 () and P2 ()

a given delay pattern), total time to complete two speech-hearing cycles, and averages and standard deviations of meter segment times.

One can, of course, generate numerous additional examples of patterns and their possible effects on rhythm, timing, etc. The question of practical importance to the design and operation of real-world systems is whether or not any of the phenomena that can be demonstrated on paper have a demonstrable impact on the quality of conference proceedings, on the ability to solve problems accurately and within a prescribed time, and/or on the collective effort required of conferees to use a given system effectively.

Delays introduced into communications among conference participants represent a type of constraint that must, in some sense, be "lived with." They are a consequence of doing business at a distance and they will always be with us, at least in principle. There are other types of constraints over which the designer of a

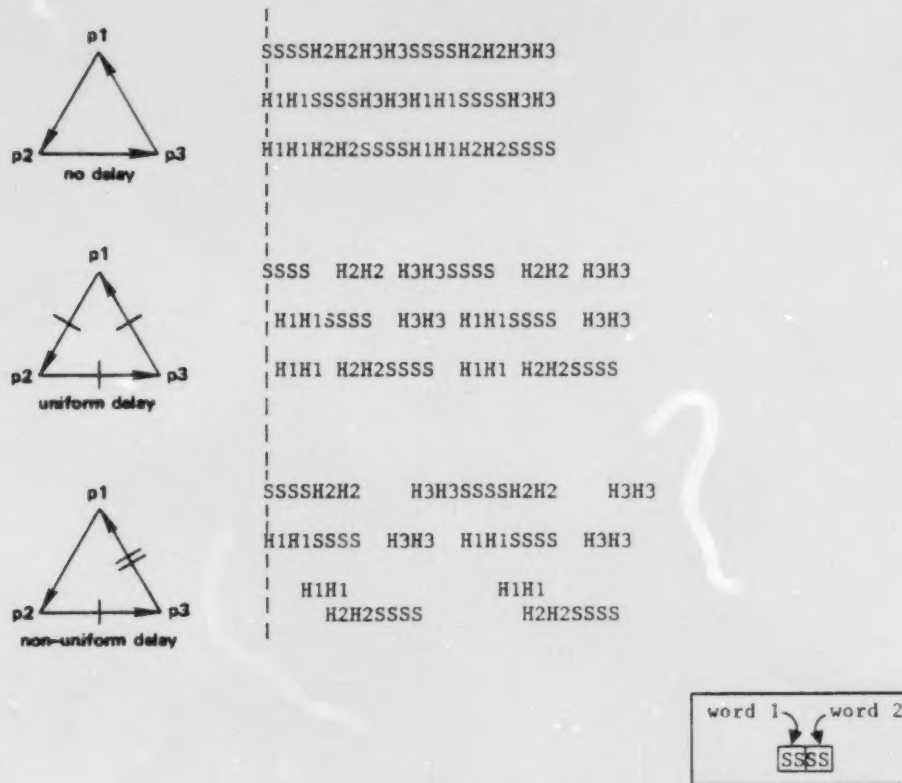


Fig. 4—Summary of events during completion of a simple communication task in systems with delay

communications system can exercise some control. It is with respect to these factors that both he and the subsequent user might benefit immensely from the availability of a data base concerned with human factors in communication. The next set of figures will attempt to outline a few of what seem to be the more important issues.

Figure 6 presents a simplified picture of a speech energy waveform, showing the rise in detected energy as a speaker begins to talk, a fall as he reaches the end of a word or phrase, and a number of peaks and valleys which are unidentifiable. Now, if transmission of the speech energy along various speech links is corrupted by noise—a virtual certainty in military communications—or if the system is designed to select only one from among a number of competing speakers, the designer will need to set thresholds to aid in filtering out the noise and/or competing speech energy. Some of the questions that may need to be answered in connection with the design of these thresholds are listed at the bottom of the figure. At the moment, we have almost no guidance for the setting of these thresholds within the communications context discussed here. The next two figures may give some inkling as to why answers to the questions are important.

Figure 7 portrays the dynamics of two of the systems we have worked with during the past two years. The upper picture is of an analog bridge system similar

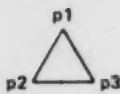
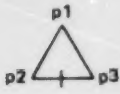
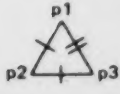
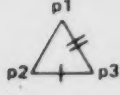
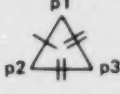
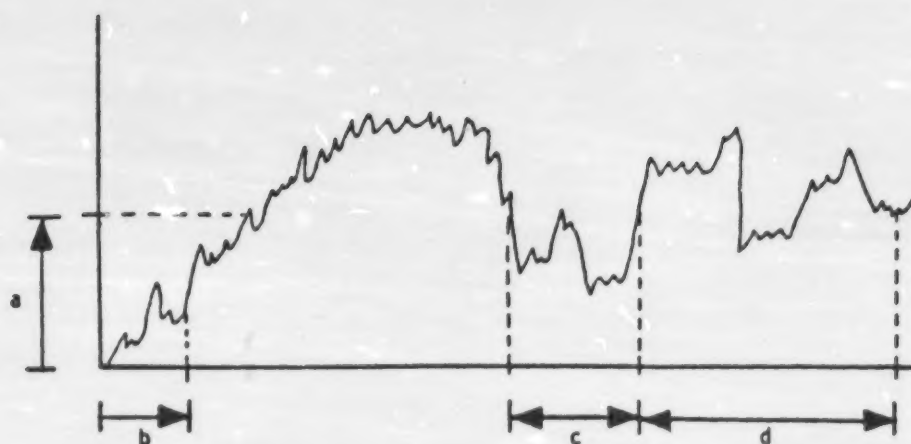
Pattern	Collisions	Meter	Total Time Complete 2 Cycles	Avg. Time Meter Segments	Std.Dev.
	none	3-3-3-3-3	24	3	0
	1 minor	3-4-3-3-4	27	3.4	0.49
	none	4-4-4-4-4	30	4	0
	2 minor	4-4-6-4-4	34	4.4	0.80
	2 major	3-4-6-3-4	32	4	1.91
	none	4-6-6-4-6	38	5.2	0.98
	none	6-6-6-6-6	42	6	0

Fig. 5—Summary of collisions and meters for various types of network patterns

to that used in normal telephone communications. With this system, when two (or more) people speak at the same time, the listener hears the sum of their inputs, i.e., hears one speaker's voice overlaid by another's. A speech energy detector (SAD), if used, helps to ensure that only the speech energy is transmitted. Depending on the setting of the SAD, more or less of the energy output of the two speakers is lost, and, as depicted in the figure, the time at which the listener is aware of a transition from the sum of speakers to Speaker 2 alone varies.

The lower portion of Fig. 7 shows the same idealized wave forms in the context of a "Simplex Broadcast" system. Unlike the "bridge," this system selects the voice of only one speaker in the event that there is a collision or interruption, and listeners hear only that selected voice. When the energy in the selected voice falls below threshold, the voice of a second speaker can then be heard. If we recall the collision and interruption figures shown earlier in the context of this and Fig. 8, we



if I want to increase the certainty that energy detected on a channel is due to speech

a: Do I set an energy threshold that must be exceeded before transmission can take place?

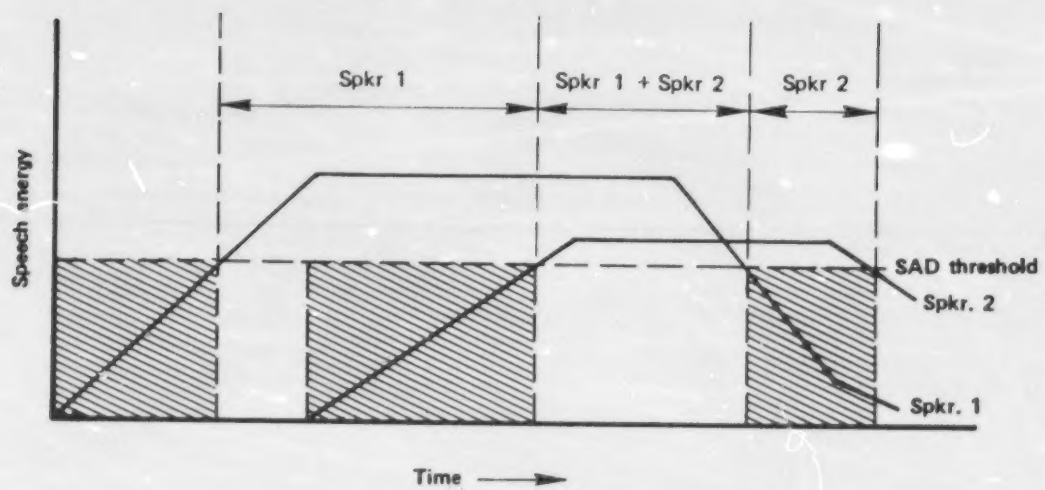
OR b: Do I set a temporal threshold, requiring that energy be present for some period of time before transmission?

OR a,b: Do I use a combination of thresholds? Or a slope criterion? Etc.

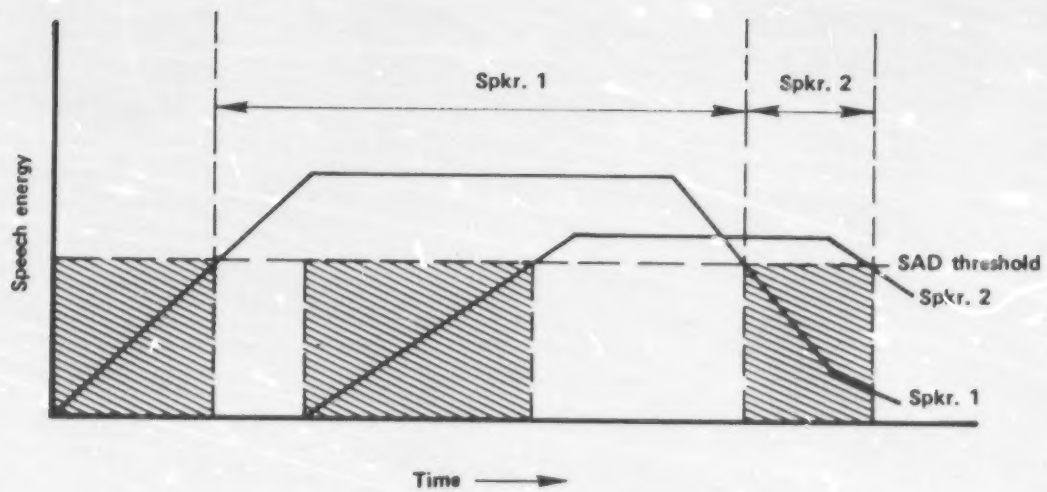
c: Is the speaker finished or has he simply paused?

d: Are these speech sounds? Speech plus noise? Articulation sounds? In any case, do I want to transmit them?

Fig. 6—A few of the questions that may need to be answered in the design of a voice communication system



Analog Bridge dynamics



Simplex Broadcast dynamics

Fig. 7—Analog bridge and simplex broadcast dynamics

can begin to imagine what difficulties listeners to a conversation using the Simplex Broadcast might have.

Figure 8 presents the dynamics of two other leading contenders for voice conferencing systems. In the upper portion of the figure, listeners hear only the voice of the selected speaker until its energy falls below threshold. They then hear the second speaker, as before. In this form of "Speaker-Interrupter" system, whoever is chosen to be speaker hears the voice of the interrupter, while the former holds the input channel to the conference. One argument in favor of this system is that if a given speaker can hear someone else attempting to interrupt him, he might be less likely to "camp" on the line. A possible problem with this and the following system in some circumstances is that, since both speaker and interrupter can hear each other, they may have difficulty determining whose voice is being transmitted to the rest of the conference.

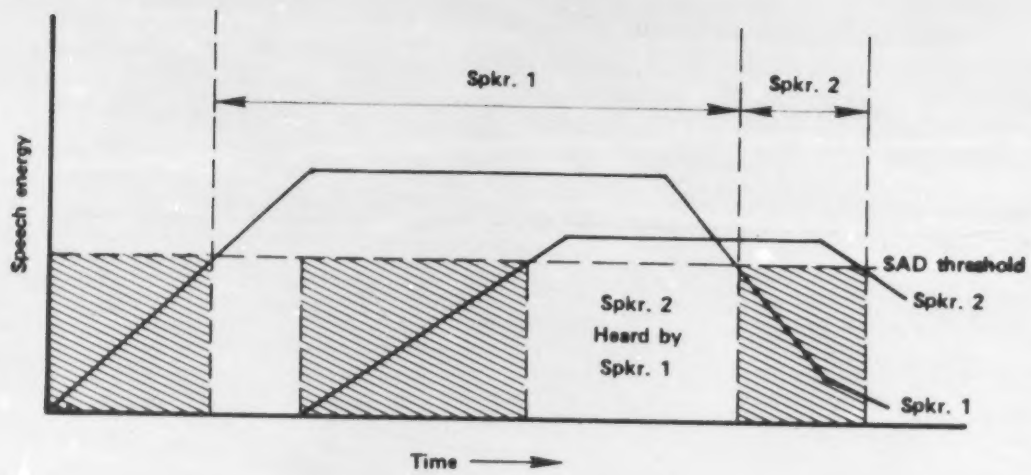
Another version of the Speaker-Interrupter system is shown in the lower portion of Fig. 8. Here, the system transmits the voice of the first speaker over the threshold, adds the voice of an interrupter (Speaker 2), and then continues with the latter as Speaker 1's energy drops below threshold. This arrangement maintains some of the advantages of a normal bridge: Listeners are aware that more than one person is attempting to talk and may be able to gain more information. The overlay of voices may, however, present a problem, as it occasionally does in the bridge. And, of course, there is no guarantee that, given the context of the discussion, the interrupter whose voice was selected was necessarily the more desirable one for the conference to hear.

Figure 9 presents a summary of what listeners hear in four different conflict situations in the four conferencing systems just described. I have included another outcome, "or nothing," in some cells of this table, since, in some configurations of these systems, nothing is heard when two or more speakers are in conflict.

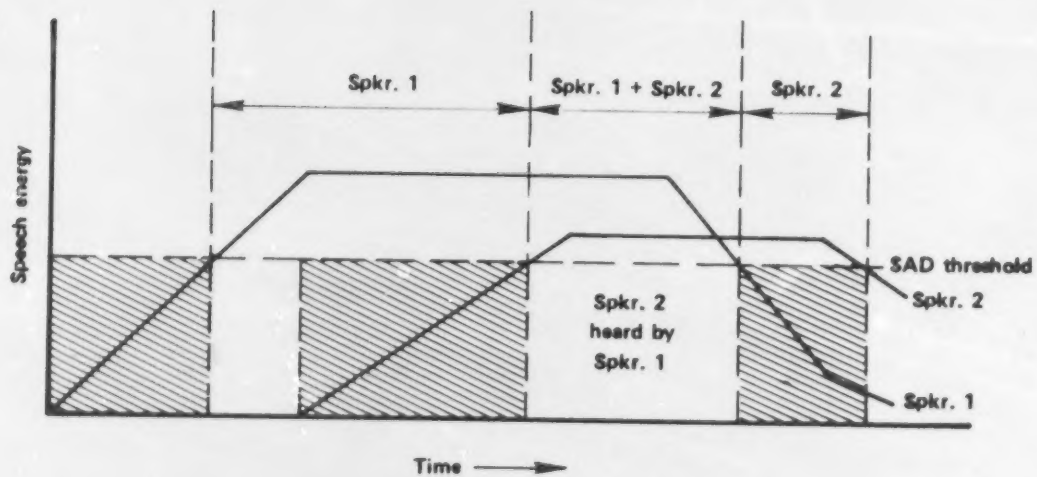
There are a number of other factors that, singly or in the aggregate, present an impediment to the free flow of conversation during a conference. I would like to mention these and then get on to a discussion of the research that we are performing. Table 1 lists the major system and conferencing variables with which we have been concerned in our studies up to this point. Note that the list includes such additional variables as control (that is, whether the conference contains a chairperson), speech vocoding (a requirement in most secure voice links), conference size and type, and control locus. Although we have just begun to do research on the distributed control systems, we are already convinced that they may provide one of the most difficult conferencing environments to be evaluated.

Now, before summarizing some research needs in this area, I want to say a few words about the methods and procedures we have used in our first two years of study.

We discovered very early in our work that the systems we were evaluating could be differentiated only in circumstances where there was a significant amount of competition for conference channels, that is, where reasonably high frequencies of collision and interruption occurred. Accordingly, we set out to develop conference tasks that, by their nature, created conflicts among participants. Table 2 summarizes the types of tasks and measures used in the evaluation. Examples of the actual tasks used include simple word-response tasks, where each participant must respond as quickly as possible to a cue word output by a previous speaker;



Speaker-Interrupter dynamics: Type I
(Interrupter channel to Speaker 1)



Speaker-Interrupter dynamics: Type II
(Interrupter channel to Speaker 1 and listeners)

Fig. 8—Speaker-interrupter dynamics

EVENT PATTERNS

LISTENERS HEAR

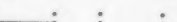
1. <u>Simultaneous Speech</u>	Bridge	Simplex Broadcast	Speaker & Inter. to Speaker	Speaker & Inter. to Speaker & Listeners
Spkr. A.  Spkr. B.  Spkr. C. 	A +B +C	A or B or C or nothing	A or B or C or nothing	or $\begin{bmatrix} A \\ B \end{bmatrix}$ or $\begin{bmatrix} B \\ C \end{bmatrix}$ or $\begin{bmatrix} A \\ C \end{bmatrix}$ or nothing
2. <u>Interruption</u>				
Spkr. A.  Spkr. B.  Spkr. C. 	A + B + C	A or A _s	A or A _s	A + B
3. <u>Collision</u>				
Spkr. A.  Spkr. B.  Spkr. C. 	A A +B +B +C +C	A A or B or B or C or C or nothing	A A or B or B or C or C or nothing	$\begin{bmatrix} A \\ B \end{bmatrix}$ or $\begin{bmatrix} A \\ C \end{bmatrix}$ or $\begin{bmatrix} B \\ C \end{bmatrix}$ or nothing
4. <u>Clear Speech</u>				
Spkr. A.  Spkr. B.  Spkr. C. 	A + B + C	A + B + C	A + B + C	A + B + C

Fig. 9—Conference speech events and what listeners hear in four different communication systems

Table 1

TYPES OF TELECONFERENCING VARIABLES STUDIED

- o Speaker Selection
 - o Uncontrolled
 - o Controlled
 - o Via Chairperson
 - o Via computer, using touch tone or push button signal
 - o Via computer, using voice energy
 - o on basis of priority
 - o on basis of loudest voice
 - o on basis of first speaker detected
- o Transmission of Voice of Interrupter
 - o To no one
 - o To speaker
 - o To speaker and listeners
- o Type of Voice Encoding
 - o No encoding
 - o EVSD (continuously variable slope delta)
 - o LPC (linear predictive coding)
 - o APC (adaptive predictive coding)
- o Number of Participants
 - o Small conferences: 4-8 participants
 - o Large conferences: 9-20 participants
- o Type of Conference
 - o Information dissemination
 - o Problem solving
- o Satellite Delay
- o Control Locus
 - o Central
 - o Distributed

Table 2

METHODS OF STUDY

-
1. Simple Conference Tasks Requiring Information Dissemination and Retrieval.
 - o Times taken to complete tasks
 - o Error rates
 - o Opinions of Conferees
 2. Complex Problem Solving Tasks Requiring Information Dissemination and Retrieval, Inference and Decision Making.
 - o Times taken to complete tasks
 - o Error rates
 - o Quality of solutions
 - o Collision and interruption measures
 - o Opinions of conferees
-

a consensus task in which participants must agree on the distribution of qualities and traits that make up the ideal lawyer (for example); and a more complex "tele-war" military planning task, in which participants are divided into a chairperson, planners, and staff, each category of participant having different partial information which must be integrated and formulated into an effective plan.¹

In Table 3 I have listed some of the more prominent needs that we see as a result of our work. Research in some of these areas is being conducted but, in my judgment, we are far from having data bases sufficient for identifying types of conferencing systems that will work efficiently and well, and for developing team training requirements and protocols in this sort of environment.

¹ For brevity, detailed discussions of scenario tasks and related figures have been omitted in this paper.

Table 3

SOME RESEARCH NEEDS

o Empirical Studies

- o to characterize normal pause, pause-in-conflict, articulation, etc. durations for a wide population of users
- o to determine the marginal gains possible with the addition of video and/or hardcopy channels to voice links
- o to determine the value of "virtual conference" displays
- o to determine what aids a chairperson might usefully employ

o Testing Paradigms and Performance Measures

- o for creating realistic interruption and collision situations in which participant interactions can be closely monitored
- o for assessing individual and team effort required in the use of conferencing systems

o Analytic and/or Simulation Models

- o number of participants in conference
- o connectivity of participants
- o bandwidths of communication channels serving conference
- o system-induced delays
- o user-induced delays

o Training Routines and Aids

THE ROLE OF HUMAN FACTORS ENGINEERING IN TEAM PERFORMANCE RESEARCH

David Meister

The purpose of this paper is to describe how human factors engineering (HFE) relates to team performance.

HFE is the application of behavioral principles and data to the design and development of man-machine systems (MMS). HFE activities begin in the very early planning stages of MMS development, when almost all critical personnel-related decisions are made, and extend to the point at which the developed system is handed over to its users. Beyond that time, HFE may be utilized when personnel-related problems arise in the operation and maintenance of the system or if, as often occurs, the system must be modified.

During the system development period, which may last as long as three or four years for complex systems, the HFE engineer performs such activities as function/task analyses in the early planning stages, which lead to the determination of the number and types of personnel required, the functions they perform, and specification and description of the tasks to be performed. Later in development, he helps design work stations and develops and/or reviews operating and maintenance procedures. Still later he helps to evaluate whether the system is performing according to requirements.

One may ask what this has to do with team performance. The answer is that, if a team exists, the specific dimensions that team will have are determined by the system configuration which the HFE engineer helps to create. Those who later have the responsibility for developing team training curricula or who attempt to evaluate the adequacy of team performance start with the team *as a given*, as already in existence. The HFE engineer makes no such assumption; it is his responsibility to develop a team *if a team is needed to exercise the system efficiently*.

One will never encounter a system specification that requires the developer to construct his system using teams or, alternatively, single operators. In fact, few system specifications include any personnel-related requirements at all except the time-honored one of making the system operable by incorporating "good" HFE principles into its design.

In consequence, the development of a team is a side effect of design processes that do not even consider the question of teams versus single operators. This makes it necessary to do two things: first, to ask whether teams should be required at all, and next, to examine the design/development process to see why teams actually become a part of the system.

If one looks dispassionately at what is involved in team performance, it is immediately apparent that teams are a poor solution to the problem of making a system function.

First, a team is expensive because it presumably requires training over and above the individual training given to individuals in the team. We will not even consider the issue of whether team training as such assists team performance more than would an equivalent amount of additional operator training; experimental

research (see Meister, 1976) has not yet been able to answer that question. If, however, we take a real-world example such as a football team, it is apparent that to make a football team mesh as a unit requires extensive training over and above that required to make a single football player proficient. Our own Navy studies tell us that training costs are a major component of system life-cycle cost. It may be true that the Navy expends little formal training time on team training, leaving it to the on-the-job situation; but if it did provide formal team training (as perhaps it should), that training would be very expensive. From that standpoint the fewer teams, the better.

Second, the probability of error and the likelihood of failure to perform a task increases when teams must be exercised. A team relationship is a dependency relationship between at least two operators; and it can be shown mathematically that the probability of successful task completion (P_t) is reduced by the inclusion of each additional operator in the team. For example, if P_{t1} (operator 1) is .9999 and P_{t2} (second operator) is also .9999, the combined resultant P_t is .9998—obviously, something less than each operator's P_t . The reason is that the team dimension (whatever it is) imposes a skill demand on personnel that makes error occurrence or failure to satisfy the demand more likely. (The reduction in P_t occurs of course only in series-dependent tasks, i.e., where every task in a job must be completed satisfactorily or the entire job is failed. The effect of independent or parallel tasks on team task performance probability is more complicated, but the general principle that each new team member reduces the likelihood of task performance by some amount still holds.) I must emphasize that it is the interaction between operators that reduces P_t ; if a team member acts independently and does not interact with other team members, there is no P_t loss and there may even be a gain.

The HFE engineer's primary responsibility in system development is to reduce skill demands wherever possible and thus to increase the likelihood of task completion and satisfactory system performance. From that standpoint, teams are "bad" and ways should be sought to avoid them. For most team performance researchers, the notion that teams may be good or bad for system functioning may seem irrelevant because the researcher accepts the team as a given, but to the engineer involved in system development, given the choice between single operators and teams, the former may be preferable.

As a member of the system development team, the HFE engineer represents what might be termed the "user" of team performance research. There are several users who have an interest in research outputs: first, system personnel themselves—reducing team demands makes it easier for them to do their job; second, system engineers who attempt to develop the most efficient systems they can—it is a reasonable hypothesis that systems depending on teams are probably less efficient than those composed of single operators; and, third, HFE engineers assisting design engineers in system development who need practical guidelines on how to apply behavioral research results to system development. Any proposed research on teams should anticipate the kinds of conclusions that will be derived from that research and particularly how these conclusions can be applied to system development. If these conclusions cannot be applied to system development, they will be of limited value to Navy planners.

If the Navy is concerned about team performance, it is inadmissible merely to accept the existence of teams as a by-product of the design process—which is the

case at present. To do so is merely to say we will accept whatever the system developer gives us, *whether or not it is an optimal answer to system requirements*. It is incumbent upon system planners for whom we do this research (among others) to ask themselves, Do we want to emphasize or de-emphasize the use of teams in future system development? Is the *creation* of teams, as distinct from their training, out of our hands? Design is a deliberate and deliberative procedure. If the developer is asked to make his system function around single operators or teams (and given that he is provided with the techniques to do so), he will attempt to satisfy the requirement.

One cannot of course create by fiat a system solely of teams of single operators, because inherent system requirements may still demand more of one or the other, and these inherent requirements cannot be overridden. However, one can think in terms of *degrees* of "teamness" so that by proper manipulation of design one might increase or reduce the use of teams.

To manipulate "teamness" requires a precise definition of the dimensions that make up a team performing specific functions and tasks. (This definition is needed also by the instructional system developer and by those who test and evaluate team performance. One must know what one is operating on before laying scalpel to flesh.)

If teams and their characteristics are to be manipulated during system development, very specific conclusions will be required. One cannot say to the engineer "design your system in terms of teams" or "design the system to make use of single operators only"; he does not know what to do with such a requirement. Those requirements written into system specifications for which no implementing techniques are provided will usually be ignored because the engineer does not know how to deal with them.

The first requirement of team performance research therefore is to define what makes up "teamness" in terms of dimensions that can be identified objectively from such system elements as task analyses or operating procedures.

It may appear as if I am returning to basics when I say that the starting point in team performance research is the identification of those dimensions that make up a team. However, just about the only dimensions consistently reported in the literature are interaction and communication; these do not appear to be a sufficient, however necessary, set of dimensions. Having identified what makes a team, it then becomes necessary to link these factors or dimensions to the design concepts and processes used by the engineer.

Although it is obvious that it is necessary to define the qualities that make up a team (as opposed to a single operator), it is equally important that these dimensions be defined in terms that anyone and particularly the system developer and HFE engineer, can use. It is, for example, no good doing an elaborate factor analysis of a series of tests given to teams and identifying Factor (I), interaction, Factor (II), communication, Factor (III), feedback, without specifying how one can identify those factors in reviewing step-by-step operating instructions, for example, and how one can make use of the resultant information.

Comparatively little is known about how the engineer designs; the problem is a very difficult one and only limited research has been performed (e.g., Meister, 1971; Askren and Lintz, 1975). Indeed it is very difficult even to conceptualize the concepts an engineer incorporates into his design (although, see Mackie et al., 1979,

on this point). In any event, if we wish systems to emphasize or de-emphasize teams, it will be necessary to show the engineer how to do so, because he does not know how. In the event that there is some question about whether it is possible through design to affect the team and its characteristics, I have attempted to list the ways in which design can impact upon team constituents and performance (this occurs primarily through the allocation and distribution of functions decided on early in design):

1. Changing the *number of personnel* who must work together and who thus must be integrated into a team.
2. Changing the *amount and type of interaction* within the team. This is what is meant by "teamness." For example, it is possible to have only two system operators interact only once in a mission; or one could have four or five operators interacting constantly throughout the mission. The latter can be considered more of a team than the former. Moreover, the interaction will be heavily influenced by the nature of the task; whether the task is psychomotor, communicative, or cognitive, will have, I suspect, different connotations for team performance.
3. Changing the *allocation of functions* within a team, e.g., taking a function from one operator and giving it to another. Or a function may be automated, which means that it is no longer team-related.
4. Changing the *amount and type of information* given individuals within the team and the team as a whole. This is the question of feedback to the team. Teams, like individuals, do not improve their performance unless they receive information about the consequences of their actions. I shall return to this point later.
5. Changing the *length of time* over which individuals must function as a team.
6. Changing the *organization* of the team as part of the larger system organization, e.g., affecting the amount of interaction between one team and another or between subsystems.

At this point, perhaps one should ask what information about teams the user would like to have and then see how this information might be used during system development.

1. How does one *recognize* a set of conditions during system development that leads to a team requirement? The development engineer needs to know when he must apply design processes to these conditions. The answer depends, of course, on the definition of team dimensions already referred to.
2. What can and should the engineer do to his design configuration in order to increase or decrease "teamness"? This question relates to the linkage between team dimensions and design processes already referred to.
3. How does one *organize* a team in order to maximize its efficiency? Operationally, this means, How should one develop operating procedures to reduce the probability of error in team performance and the time needed to reach a desired level of team performance?
4. How does one design work stations and information displays to permit the

team to function most effectively? What kind and amount of information and information feedback does a team need in order to be efficient?

5. How does one develop a training program to develop team qualities apart from the learning required to develop single operator skills? The answer to this question is properly the province of the instructional system developer, but it depends, as the previous questions also depend, on the definition of team dimensions. If one does not know what one is supposed to train, training is hardly possible.
6. How does one measure the efficiency of a team as a team, apart from single operator measures? From a system evaluation standpoint, the only important question is whether the system operates (or does not operate) according to requirements. Thus, only system outputs (composed of functions performed by teams, individuals, and subsystems) matter. However, when the system fails to operate according to requirements, it is necessary to look at how effectively the team has performed, and this makes it necessary to have methods of examining team effectiveness.

Answers to these questions are needed at various phases during system development. For example, it is possible to emphasize or de-emphasize teams by actions taken during the function/task analyses (such as those required by MIL-H-46855B (1979)). In the performance of these analyses it is possible to shift functions from humans to equipment and vice versa and to change the ways in which functions are to be performed—all of which affects the team. What we need to know, however, is what specific effects such reallocations will have on "teamness." For example, how much reduction in interaction requirements will produce how much effect (and what effect) on team efficiency? At this stage of system development, the designer can be influenced because alternative configurations are still viable possibilities; design has not yet been "locked in."

At a somewhat later stage of function/task analysis, tasks are devised to implement functions, and these (tasks) can now be examined to determine, for example, that interaction requirements among operators are not excessive; that is, not too much must be done in too short a time to meet too rigorous a standard of precision. At this stage, we must know more precisely what team performance dimensions are so that they can be identified from examination of operating procedures.

The design of the work station during detail design to optimize operator interactions is not too difficult if again team dimensions can be identified from operating procedures.

A special problem arises with the determination of feedback information requirements to maximize team performance. Information needed to perform the individual operator's job is not difficult to determine. However, *feedback information* relevant to the team as a team is a different story, because several different types of feedback are possible: (1) information about the performance of the individual operator; (2) information about the performance of the team as a team; (3) information about the performance of the system to which team performance is an input. Information about team/system performance is probably less relevant to the individual operator than information about his own performance. On the other hand, some team feedback information (even if incorrect) is necessary if the team is to improve (see Meister, 1976, for a more detailed discussion of these points).

Finally, if system performance is inadequate and if that system is composed (in part) of teams, it is necessary to evaluate the adequacy of the team performance in terms of known team dimensions. Again, it is necessary to revert to the question we began the discussion with: What are the dimensions that define a team?

Here it is necessary to issue a warning which also becomes a research requirement. The team dimensions that are our starting point for analysis and evaluation may vary somewhat as a function of the type of task/job being performed. It is entirely conceivable that interaction (if that is a team dimension) is somewhat different in its effects on cognitive as differentiated from psychomotor tasks. Team dimensions may be more or less specific to tasks; we do not know whether this is true, but we must try to find out by investigating a spectrum of jobs in which teams are involved.

Let me conclude by considering the research methods available to answer the questions posed by HFE in team performance. Although HFE is a discipline somewhat independent of psychology, it makes use of all the techniques available to behavioral science: experimentation, simulation, mathematical modeling, field studies, content analysis of documents, etc. One of its research methods which is not often used but which I recommend is what can be called the "study of system history."

The Navy, the other military services, civilian industry, and the world about us build MMSs in great number, in the course of which teams are developed, trained, and tested. It is my impression that there is very little systematic recording of data describing what happens during the development process; if data are indeed recorded, almost no one examines them. We fail then to use our available historical resources to answer questions such as:

1. What was the conceptual and design process by means of which the team requirement was first identified?
2. What was the rationale (i.e., function allocations) upon which a particular team configuration was decided?
3. How was planning of team training conducted?
4. During system development, what tradeoffs were made that impacted on the team, its organization, utilization, etc.?
5. How was team performance tested (if it was)?
6. What changes occurred during a particular system development that influenced the development of teams and their characteristics?

Such data are not easy to secure, but ONR might consider funding (on a multi-year basis, of course) an organization which would attempt to monitor a limited number of Navy systems during their development by interviewing the participants and reading pertinent documents to build up case histories that would illuminate system development from the standpoint of team performance. I do not suggest that this should be the sole methodology employed; rather, it should be used to supplement more traditional research methods such as experimentation. Such an investigative approach might suggest which research solutions can or cannot be applied to the operational environment.

Let me summarize the tenor of my remarks. These focus on:

1. Concern for the user. Research that does not meet his needs is useless except to the researchers who create that research.

2. The necessity for *doing something* about teams in system development rather than merely accepting their existence as developmental emergents.
3. The necessity for *defining the dimensions* that make up teams and for interrelating these dimensions with the engineer's design concepts.
4. The necessity for giving the system developer *tools* to make use of team performance research.
5. The necessity for *concentrating on questions* that will be asked by users during the course of system development.
6. The desirability of using an *historical approach* as a supplemental research technique and as a means of testing the adequacy of the research solutions proposed.

HFE is user-oriented and activist in its research strategy; it is applied in the research answers it seeks; it is utilitarian in its evaluation of research success. One would hope that a program of team research would be similarly oriented.

DISCUSSION OF SESSION 8

Feehrer began the discussion with a plea for broad-based research on teams that considers both human and machine components and their interactions. Furthermore, research should address the issue of interaction among machines; in the future, we may see entire fleets of unmanned submarines that must communicate and coordinate their actions.

Rick Hayes-Roth noted that the complexity of a task varies greatly with the requisite equipment. Technology creates both opportunities and constraints; human factors research should exploit the former and adapt to the latter. As an example, he pointed out that complex displays contributed to the emergency in the Three Mile Island nuclear accident. He stressed the need for intelligent selection of information to be displayed in such high-stress environments.

Thorndyke pointed out that the word exchange tasks used in Feehrer's teleconferencing research required little cognitive activity in the interim between messages. The major bottleneck in this situation is the rate of information transfer. This condition will not necessarily obtain for more complex naturalistic tasks. Feehrer replied that the word tasks gave comparable results to more complex tasks only because the dependent measures were very crude; namely, the total time to complete the task. Interesting differences between tasks would only be captured by using measures sensitive to semantic content and organization. However, the word tasks do satisfy research requirements for tasks necessitating quick and accurate responses.

Edwards asked rhetorically why humans should be included in a human-machine system at all. The lack of effective hardware and the relative economy of humans versus machines will probably vanish as reasons over the next few decades, as technology becomes increasingly sophisticated. The remaining reason is that there are certain tasks, such as deciding when to fire nuclear missiles, that one would not want a machine to perform even if it were technologically and economically feasible. Edwards suggested that research should examine such situations in which humans are not expendable.

Rick Hayes-Roth noted that every increase in automation increases the complexity of the system and hence the difficulty of the human's task. Edwards, in contrast, proposed that the primary human role in a human-machine system is to operate the on/off switch.

Parsons injected a final comment on estimating the rate of automation. He noted that predicted rates of progress are often much greater than actual rates. As a case in point, he noted that the first teleconferencing studies were conducted in 1951, yet we are still grappling with many of the same issues. In short, the world does not change as rapidly as one might think.

BIBLIOGRAPHY

- Allen, J. F., *A Plan-Based Approach to Speech Act Recognition*, Tech. Rep. No. 131/79, University of Toronto, Dept. of Computer Science (Doctoral dissertation), 1979.
- Allen, R.F., and S. Silverzweig, "Group Norms: Their influence on Training Effectiveness," in R. L. Craig (ed.), *Training and Development Handbook*, McGraw-Hill, New York, 1976.
- Allison, G., *Conceptual Models and the Cuban Missile Crisis: Rational Policy, Organization Process, and Bureaucratic Politics*, (Doctoral thesis), The Rand Corporation, P-3919, August 1968.
- , *Essence of Decision*, Little-Brown, Boston, 1971.
- Ammerman, H. L., and F. C. Pratzner, "Occupational Survey Report on Automotive Mechanics: Task Data From Workers and Supervisors Indicating Job Relevance and Training Criticalness," R & D Series No. 110, Center for Vocational Education, Columbus, Ohio, 1975.
- , "Occupational Survey Report on Business Data Programmers: Task Data From Workers and Supervisors Indicating Job Relevance and Training Criticalness," R & D Series No. 108, Center for Vocational Education, Columbus, Ohio, 1974.
- , *Performance Content for Job Training*, Vols. 1-5, R & D Series No. 121-125, Center for Vocational Education, Columbus, Ohio, 1977.
- Asch, S. E., "Effects of Group Pressure on the Modification and Distortion of Judgments," in H. Geutskow (ed.), *Groups, Leadership, and Men*, Carnegie, Pittsburgh, Pennsylvania, 1951.
- , "Studies of Independence and Conformities," *Psychological Monographs* Vol. 70, No. 9(Whole No. 416), 1956.
- Askren, W. B., and L. M. Lintz, "Human Resources Data in System Design Trade Studies," *Human Factors*, Vol. 1, No. 17, 1975, pp. 4-12.
- Austin, J. L., *How To Do Things With Words*, Oxford University Press, London, 1962.
- Baumgartel, H., and P. Jeanpierre, "Applying New Knowledge in the Backhome Setting: A Study of Indian Managers' Adoptive Efforts," *Journal of Applied Behavioral Science*, Vol. 8, No. 6, 1972, pp. 674-694.
- Bentler, P. M., "Multivariate Analysis With Latent Variables: Causal Modeling," *Annual Review of Psychology*, 1980.
- Bentler, P. M., and G. Speckart., "Models of Attitude-Behavior Relationships," *Psychological Review*, Vol. 86, 1979, pp. 452-464.
- Billings, R. S., R. J. Klimoski, and J. A. Breaugh, "The Impact of Change in Technology on Job Characteristics: A Quasi Experiment," *Administrative Science Quarterly*, Vol. 22, 1977, pp. 318-339.
- Bons, T. M., and F. E. Fiedler, "Changes In Organizational Leadership and the Behavior of Relationship and Task Motivated Leaders," *Administrative Sciences Quarterly*, Vol. 21, 1976, pp. 433-472.

- Brock, T. C., "Communication Discrepancy and Intent to Persuade as Determinants of Counterargument Production," *Journal of Experimental Social Psychology*, Vol. 13, 1967, pp. 315-332.
- Brown, J. S., A. Collins and G. Harris, "Artificial Intelligence and Learning Strategies," in H. F. O'Neil (ed.), *Learning Strategies*, Academic Press, New York, 1978.
- Bruce, B. C., and D. Newman, "Interacting Plans," *Cognitive Science*, Vol. 2, 1978, pp. 195-233.
- Burnstein, E., et al., "The Risky Shift is Eminently Rational," *Journal of Personality and Social Psychology*, Vol. 20, 1971, pp. 462-471.
- Burnstein, E., and A. Vinokur, "What A Person Thinks Upon Learning He Has Chosen Differently From Others: Nice Evidence for the Persuasive-Arguments Explanation of Choice Shifts," *Journal of Experimental Social Psychology*, Vol. 13, 1977, pp. 315-332.
- Burnstein, E., A. Vinokur and Y. Trope, "Interpersonal Comparison Versus Persuasive Argumentation: A More Direct Test of Alternative Explanations for Group Induced Shifts in Individual Choice," *Journal of Experimental Social Psychology*, Vol. 13, 1977, pp. 315-332.
- Campbell, D. T., "Qualitative Knowing in Action Research," paper presented at 82nd Annual Meeting of the American Psychological Association, New Orleans, 1974.
- Campbell, J. P., "On the Nature of Organizational Effectiveness," in P. S. Goodman and J. M. Pennings (eds.), *New Perspectives on Organizational Effectiveness*, Jossey-Bass, San Francisco, 1977, pp. 13-55.
- , "What We Are About: An Inquiry Into the Self Concept of Industrial and Organizational Psychology," presidential address to the Division of Industrial and Organizational Psychology at the 86th Annual Meeting of the American Psychological Association, Toronto, 1978.
- Clark, H. H., "Responding to Indirect Speech Acts," *Cognitive Psychology*, Vol. 11, 1979, pp. 430-477.
- Cohen, P. R., *On Knowing What to Say: Planning Speech Acts*, Tech. Rep. No. 118, University of Toronto, Dept. of Computer Science (Doctoral dissertation), 1978.
- Cohen, P. R. and C. R. Perrault, "Elements of A Plan-Based Theory of Speech Acts," *Cognitive Science*, Vol. 3, 1979, pp. 177-212.
- Cole, R., and J. Morgan, *Syntax and Semantics, 3: Speech Acts*, Academic Press, New York, 1975.
- Coleman, J. S., et al., *Equality of Educational Opportunity*, Report of the Office of Education, U.S. Department of Health, Education, and Welfare, U. S. Government Printing Office, Washington, D.C., 1966.
- Collins, A., "Processes in Acquiring Knowledge," in R. C. Anderson, R. J. Spiro, and W. E. Montague (eds.), *Schooling and the Acquisition of Knowledge*, Lawrence Erlbaum Associates, Hillsdale, New Jersey, 1977.
- Collins, A. and A. L. Stevens, "Goals and Strategies of Effective Teachers," in R. Glaser (ed.), *Advances In Instructional Psychology*, Vol. 2, Lawrence Erlbaum Associates, Hillsdale, New Jersey, (in press).
- Collins, J. J., *A Study of Potential Contributions of Small Group Behavior Research to Team Training Technology Development*, ONR #N0014-76-C-1076, Essex Corporation, Alexandria, Virginia, 1977.

- Compact Edition of the Oxford English Dictionary*, 2 vols., Oxford, England, 1971.
- Corkill, D. D., "Hierarchical Planning In A Distributed Environment," in *Proceedings of the Sixth International Joint Conference on Artificial Intelligence* (2 vols.), Tokyo, August 1979, pp. 168-175.
- Fiedler, F. E., M. Chemers and L. Mahar, *Improving Leadership Effectiveness: The Leader-Match Concept*, Wiley, New York, 1976.
- Fiedler, F. E., et al., "Organizational Stress and the Use and Misuse of Managerial Intelligence and Experience," *Journal of Applied Psychology*, 1979.
- Fischer, and P. Larkey, "Organization Theory and Military Threat Assessment," paper presented at the Meetings of The Association for Public Policy Management, Chicago, Illinois, 1979.
- Flanagan, J. C., "Education: How and for What?" *American Psychologist*, Vol. 28, 1973, pp. 551-556.
- Freeberg, N. E., "Criterion Measures for Youth-Work Training Programs: The Development of Relevant Performance Dimensions," *Journal of Applied Psychology*, Vol. 61, No. 5, 1976, pp. 537-545.
- Gilbert, T. F., "On The Relevance of Laboratory Investigation of Learning to Self-Instructional Programming," in A. A. Lumsdaine and R. Glaser (eds.), *Teaching Machines and Programmed Learning*, National Education Association, Washington, D. C., 1960.
- Glaser, R. and A. J. Nitko, "Measurement in Learning and Instruction," in R. L. Thorndike (ed.), *Educational Measurement*, 2nd ed., American Council on Education, Washington, D.C., 1971.
- Goldstein, I. L., "The Pursuit of Validity in the Evaluation of Training Programs," *Human Factors*, Vol. 20, 1978, pp. 131-144.
- , "Training in Work Organizations," *Annual Review of Psychology*, Annual Reviews, Palo Alto, Calif., 1980, pp. 229-272.
- , *Training: Program Development and Evaluation*, Brooks/Cole, Monterey, Calif., 1974.
- Goodman, P. S. and J. M. Pennings, *New Perspectives on Organizational Effectiveness*, Jossey-Bass, San Francisco, 1977.
- Greenwald, A. G., "Cognitive Learning, Cognitive Response to Persuasion, and Attitude Change, in A. G. Greenwald, et al. (eds.), *Psychological Foundations of Attitudes*, Academic Press, New York, 1968.
- Hackman, J. R., "Changing Views of Motivation in Work Groups," paper presented at NATO Conference, Thessaloniki, Greece, August 20-24, 1979.
- Hackman, J. R. and E. E. Lawler, "Employee Reactions to Job Characteristics," *Journal of Applied Psychology*, Vol. 55, 1971, pp. 259-286.
- Hackman, J. R. and G. Oldham, "The Design of Work for Groups and Groups for Work," (in press).
- Hackman, J. R. and L. J. Suttle (eds.), *Improving Life at Work*, Goodyear, Santa Monica, Calif., 1977.
- Hall, E. R., and Rizzo, W. A., *An Assessment of U.S. Navy Tactical Team Training*, TAEG Report No. 18, Training Analysis and Evaluation Groups, Orlando, Florida, 1975.
- Hand, H. H., and J. W. Slocum, "A Longitudinal Study of the Effects of a Human Relations Training Program on Managerial Effectiveness," *Journal of Applied Psychology*, Vol. 56, No. 5, October 1972, pp. 412-417.

- Hannan, M. T., *Aggregation and Disaggregation in Sociology*, Heath, Lexington, Mass., 1971.
- Hayes-Roth, B. and F. A. Hayes-Roth, "A Cognitive Model of Planning," *Cognitive Science*, Vol. 3, 1979, pp. 275-323.
- Hewitt, C., "Viewing Control Structures As Patterns of Passing Messages," *Artificial Intelligence*, Vol. 8, No. 3, 1977, pp. 323-364.
- Hulin, C. L. and D. M. Rousseau, "Analysis of Infrequent Events," in K. H. Roberts and L. Burstein (eds.), *Aggregation Issues In The Social Sciences*, Jossey-Bass, San Francisco, 1980.
- Human Engineering Requirements for Military Systems, Equipment and Facilities*, MIL-H-46855B, January 31, 1979
- Hunt, E. B., "The Design of A Robot Mind," in M. Friedman, J. P. Das and N. O'Connor (eds.), *Intelligence and Learning*, Plenum Press, London, 1980.
- Janis, I. L. and L. Mann, *Decision Making*, Free Press, New York, 1977.
- Joreskog, K. G., "Analyzing Psychological Data by Structural Analysis of Covariance Matrices," in D. Krantz, et. al. (eds.), *Contemporary Developments in Mathematical Psychology: Measurement, Psychophysics, and Neural Information Processing*, W. H. Freeman and Co., San Francisco, 1974.
- Kanarick, A. F., D. G. Alden, and R. W. Daniels, "Decision Making and Team Training in Complex Tactical Training Systems of the Future," in V. Amico, J. Wootton and J. Regan (eds.), *NTDC 25th Anniversary Commemorative Technical Journal*, Naval Training Device Center, Orlando, Florida, 1972.
- Kaplan, M. F., "Discussion Polarization Effects in a Modified Jury Decision Paradigm: Informational Influences," *Sociometry*, Vol. 40, 1977, pp. 262-271.
- Kaplan, M. F. and C. E. Miller, "Judgments and Group Discussion: Effect of Presentation and Memory Factors on Polarization," *Sociometry*, Vol. 40, 1977, pp. 337-343.
- Keating, J. P., and T. C. Brock, "Acceptance of Persuasion and the Inhibition of Counterargumentation Under Various Distraction Tasks," *Journal of Experimental Social Psychology*, Vol. 10, 1974, pp. 301-309.
- Kilman, R. H., L. R. Pondy, and D. P. Slevin, *The Management of Organization Design*, (2 vols.), North Holland, Amsterdam, 1976.
- Klaus, D. J., and R. Glaser, *Increasing Team Proficiency Through Training*, 8. *Final Summary Report*, Office of Naval Research Psychological Services Division Personnel and Training Branch Contract Nonr-2551(00), NR-154-079, Washington, D.C., 1968.
- Kogan, N., and M. A. Wallach, *Risk Taking: A Study in Cognition and Personality*, Holt, Rinehart & Winston, New York, 1964.
- Laabs, G. J., R. C. Pannell, and E. J. Pickering, "A Personal Readiness Training Program: Maintenance of the Missile Test and Readiness Equipment (MTRE MK 7 MOD 2)," NPRDC Technical Report No. 77-19, 1977.
- Lesser, V. R., and D. D. Corkill, "The Application of Artificial Intelligence Techniques to Cooperative Distributed Processing," *Proceedings of the Sixth International Joint Conference on Artificial Intelligence*, (2 vols.), Tokyo, August 1979, pp. 537-540.
- Lesser, V. R., and L. D. Erman, *An Experiment in Distributed Interpretation*, University of Southern California Information Sciences Institute Report No.

- ISI/RR-79-76 (also, Carnegie-Mellon University Computer Science Dept. Technical Report CMU-CS-79-120), May 1979.
- Levin, J. A., and J. A. Moore, "Dialogue-Games: Metacommunication Structures for Natural Language Interaction," *Cognitive Science*, Vol. 1, 1977, pp. 395-420.
- McCall, M. W., A. M. Morrison, and R. L. Hannan, "Studies of Managerial Work: Results and Methods," Technical Report No. 9, Center for Creative Leadership, Greensboro, North Carolina, 1978.
- McCarthy, J., "Epistemological problems of Artificial Intelligence," *Proceedings of the 5th International Joint Conference on Artificial Intelligence* (2 vols.), Massachusetts Institute of Technology, Cambridge, Mass., 1977, pp. 1038-1044.
- , "First Order Theories of Individual Concepts and Propositions," in Hayes, J. E., D. Michie, and L. I. Mikulich (eds.), *Machine Intelligence 9: Machine Expertise and the Human Interface*, Ellis Horwood, Chichester, 1979.
- MacCrimmon, K. R., "Descriptive Aspects of Team Theory: Observation, Communication and Decision Heuristics in Information Systems," *Management Science*, Vol. 20, No. 10, 1974, pp. 1323-1334.
- , "Elements of Decision Makers and Environments," in W. Goldbert (ed.), *Behavioral Approaches to Modern Management I*, BAS, Gothenburg, Sweden, 1970.
- , "An Overview of Multiple Objective Decision-Making," in J. Cochrane and M. Zeleny (eds.) *Multiple Criteria Decision Making*, University of South Carolina Press, Columbia, 1973 (a), pp. 18-14.
- , "Theories of Collective Decision," UBC Working Paper, prepared for the Fourth Research Conference on Subjective Probability, Utility and Decision-Making, Rome, September 1973 (b).
- MacCrimmon, K. R., and R. N. Taylor, "Decision-Making and Problem-Solving," in M. Dunnette (ed.), *Handbook of Industrial and Organizational Psychology*, Rand-McNally, Chicago, 1976, pp. 1397-1453.
- Mackie, R. R., et al., *An Engineer's Guide to the Use of Human Resources in Electronic Systems Design*, (NPRDC Technical Note 79-8), Navy Personnel Research and Development Center, San Diego, California, 1979.
- McGhee, W. and P. W. Thayer, *Training in Business and Industry*, Wiley, New York, 1961.
- McGuire, W. J., "Inducing Resistance to Persuasion: Some Contemporary Approaches," in L. Berkowitz (ed.), *Advances in Experimental Social Psychology*, Vol. 1, Academic Press, New York, 1964.
- Marschak, J., and R. Radner, *Economic Theory of Teams*, Yale University Press, New Haven, Conn., 1972.
- Meister, David, *Behavioral Foundations of System Development*, Wiley, New York, 1976 (p. 288).
- , *Human Factors: Theory and Practice*, Wiley, New York, 1971.
- Moore, R. C., "Reasoning About Knowledge and Action," in *Proceedings of the 5th International Joint Conference on Artificial Intelligence* (2 vols.), Massachusetts Institute of Technology, Cambridge, Mass., 1977, pp. 223-227.
- , *Reasoning About Knowledge and Action*, Tech. Note 191, SRI International, Artificial Intelligence Center, Menlo Park, Calif., 1979.

- Newell, A., and H. A. Simon, *Human Problem Solving*, Prentice-Hall, Inc., 1972.
- , "GPS, A Program That Simulates Human Thought," in E. A. Feigenbaum and J. Feldman (eds.), *Computers and Thought*, McGraw Hill, New York, 1963.
- O'Reilly, C. A., and J. C. Anderson, "Organizational Communication and Decision Making: Laboratory Results versus Actual Organizational Settings," presented at the XXIV International Meeting of the Institute of Management Sciences, June 1979.
- O'Reilly, C. A., and D. Caldwell, "Informational Influence as a Determinant of Task Characteristics and Job Satisfaction," *Journal of Applied Psychology*, Vol. 64, 1979, pp. 157-165.
- Palmer, J., et. al., "Some Relationships Between Visual Information Processing and Reading," Technical Report, University of Washington Psychology Department, 1979.
- Pannell, R. C., and G. J. Laabs, "Construction of a Criterion-Referenced Diagnostic Test for An Individual Instruction Program," *Journal of Applied Psychology*, 1979.
- Parsons, H. M., "The Development and Installation of a System Training Program. The SAGE ECCM Model," Report SP-265, System Development Corporation, Santa Monica, Calif., 1960.
- , *Man-Machine System Experiments*, Johns Hopkins Press, Baltimore, 1972.
- , "Simulation of Highway Accidents," in *Proceedings of the Human Factors Society—22nd Annual Meeting*, Human Factors Society, Santa Monica, Calif., 1978.
- , "What the Navy's Anti-Air Warfare Training Program Can Learn From Air Defense System Training Experience," Memorandum Report, Office of Naval Research, Washington, D. C., 1964.
- Petty, R. E., and J. T. Cacioppo, "Forewarning, Cognitive Responding, and Resistance to Persuasion," *Journal of Personality and Social Psychology*, Vol. 35, 1977, pp. 645-655.
- Petty, R. E., G. Wells and T. C. Brock, "Distraction Can Enhance or Reduce Yielding to Propaganda: Thought Disruption Versus Effort Justification," *Journal of Personality and Social Psychology*, Vol. 34, 1976, pp. 874-884.
- Porter, L. W. and K. H. Roberts, "Organizational Communication," in M. D. Dunnette (ed.), *Handbook of Industrial and Organizational Psychology*, Rand McNally, Chicago, 1976, pp. 1553-1589.
- Roberts, K. H. and L. Burstein (eds.), *Aggregation Issues in the Social Sciences*, Jossey-Bass, San Francisco, 1980.
- Roberts, K. H., C. L. Hulin, and D. M. Rousseau, *Developing an Interdisciplinary Science of Organizations*, Jossey-Bass, San Francisco, 1978.
- Roberts, K. H., et al., "Organizational Communication and Organizational Theory: A Communication Failure?" *Human Relations*, Vol. 27, 1974, pp. 501-524.
- Robinson, W. S., "Ecological Correlations and the Behavior of Individuals," *American Sociological Review*, Vol. 15, 1950, pp. 351-357.
- Ross, L., "The Intuitive Psychologist and His Shortcomings: Distortions in the Attribution Process," in L. Berkowitz (ed.), *Advances in Experimental Social Psychology*, 1977.
- Rousseau, D. M., "Assessment of Technology in Organizations: Closed versus Open Systems Approaches," *Academy of Management Review*, Vol. 4, 1979, pp. 531-542.

- Sacerdoti, E. D., "Nonlinear Nature of Plans," *Proceedings of the Fourth International Joint Conference on Artificial Intelligence*, Tbilisi, Georgia, U.S.S.R., September 1975.
- Sagan, K., *Broca's Brain*, Random House, New York, 1979.
- Salancik, G. R. and J. Pfeffer, "A Social Information Processing Approach to Job Attitudes and Task Design," *Administrative Science Quarterly*, Vol. 23, 1978, pp. 224-253.
- Salinger, R. D., *Disincentives to Effective Employee Training and Development*, U. S. Civil Service Commission, Bureau of Training, Washington, D. C., 1973.
- Schank, R. and R. Abelson, *Scripts, Plans, Goals, and Understanding*, Lawrence Erlbaum Associates, Hillsdale, New Jersey, 1977.
- Schwab, D. P. and L. L. Cummings, "Impact of Task Scope on Employee Productivity: An Evaluation Using Expectancy Theory," *Academy of Management Review* Vol. 1, 1976, pp. 23-35.
- Searle, John R., *Speech Acts*, Cambridge University Press, Cambridge, England, 1970.
- Seaver, David A., "Assessing Probability With Multiple Individuals: Group Interaction Versus Mathematical Aggregation," Research Report, Oct. 1977-December 1978, prepared in cooperation with the Social Science Research Institute, University of Southern California, Los Angeles (Report Nos. USC-001922-3-T, SSRI-RR-78-3), Decisions and Designs Inc., McLean, Virginia, 1978.
- Simon, H., *The Science of the Artificial*, MIT Press, Cambridge, Mass., 1969.
- Slovic, P., B. Fischhoff and S. Lichtenstein, "Behavioral Decision Theory," *Annual Review of Psychology*, 1977.
- Smith, R. G., "A Framework For Distributed Problem Solving," *Proceedings of the Sixth International Joint Conference on Artificial Intelligence* (2 vols.), Tokyo, August 1979, pp. 836-841.
- Staw, B. M., and F. V. Fox, "Escalation: Some Determinants of Commitment to a Previously Chosen Course of Action," *Human Relations* Vol. 30, 1977, pp. 431-450.
- Stevens, A. L. and A. Collins, "The Goal Structure of a Socratic Tutor," *Proceedings of Association for Computing Machinery National Conference*, Seattle, Washington, 1977.
- Swezey, R. W., R. B. Pearlstein, and W. H. Ton, "Criterion-Referenced Testing: A Discussion of Theory and of Practice in the Army," ARI Technical Paper 273-A18-(1)-IR-0474-RWSW, Virginia, 1974.
- Tesser, A., "Attitude Polarization As A Function of Thought and Reality Constraints," *Journal of Research in Personality*, Vol. 10, 1976, pp. 183-194.
- Tesser, A. and M. C. Conlee, "Some Effects of Time and Thought on Attitude Polarization," *Journal of Personality and Social Psychology*, Vol. 31, 1975, pp. 262-270.
- U.S. Department of Defense, Military Standards, *Human Engineering Design Criteria for Military Systems, Equipment and Facilities*, MIL-STD-1472, DoD, Washington, D.C., 1976.
- Vinokur, A., "Cognitive and Affective Processes Influencing Risk-Taking In Groups: An Expected Utility Approach," *Journal of Personality and Social Psychology*, Vol. 20, 1971, pp. 472-486.

- Vinokur, A., and E. Burnstein, "The Effects of Partially Shared Persuasive Arguments on Group Induced Shifts: A group Problem Solving Approach," *Journal of Personality and Social Psychology*, Vol. 29, 1974, pp. 305-315.
- Vinokur, A., Y. Trope and E. Burnstein, "A Decision-Making Analysis of Persuasive Argumentation and the Choice-Shift Effect," *Journal of Experimental Social Psychology*, Vol. 11, 1975, pp. 127-148.
- Weick, K. E., "Educational Organizations as Loosely Coupled Systems," *Administrative Science Quarterly* Vol. 21, 1976, pp. 1-19.
- , "Organizational Design: Organizations as Self-Designing Systems," *Organizational Dynamics*, Autumn 1977, pp. 30-46.
- Weyhrauch, R., "Prolegomena To A Theory of Mechanized Formal Reasoning," *Artificial Reasoning*, forthcoming.
- Willerman, L., *The Psychology of Individual and Group Differences*, W. H. Freeman, San Francisco, 1979.
- Williges, B. H., S. N. Roscoe, and R. C. Williges, "Synthetic Flight Training Revisited," ARL Technical Report No. 72-21, Savoy, Illinois, 1972.
- Woodward, J., *Industrial Organizations: Theory and Practice*, Oxford, London, 1965.
- , *Management and Technology*, Her Majesty's Stationery Office, London, 1958.